

### Ergebnisse einer experimentellen Studie zur Delphi-Methode

Häder, Michael; Häder, Sabine

Veröffentlichungsversion / Published Version  
Forschungsbericht / research report

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Häder, M., & Häder, S. (1994). *Ergebnisse einer experimentellen Studie zur Delphi-Methode*. (ZUMA-Arbeitsbericht, 1994/05). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-70149>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUMA-Arbeitsbericht 94/05

**Ergebnisse einer experimentellen Studie  
zur Delphi-Methode**

*Michael Häder, Sabine Häder*

August 1994

**ZUMA**

**Quadrat B2,1**

**Postfach 12 21 55**

**68072 Mannheim**

**Telefon: (0621) 12 46 - 281**

**Telefax: (0621) 12 46 - 100**

**E-mail: haeder@zuma-mannheim.de**





## Danksagung

Zu Beginn dieses Berichts möchten wir uns bei allen an den vier Wellen unserer Testbefragung beteiligten ZUMA-Mitarbeitern bedanken. Ohne sie hätte dieses Projekt nicht zustandekommen können.

Unser Dank gilt außerdem Frau Dipl.-Volkswirtin Birgit Rimmelspacher, z.Zt. Praktikantin in der Abteilung Statistik am ZUMA, die uns sowohl mit zahlreichen Hinweisen bei der Fertigstellung des Manuskripts als auch bei den statistischen Analysen im Abschnitt 6.2. unterstützt hat.

Michael Häder und Sabine Häder



## Gliederung

1. Ausgangssituation	4
2. Erklärungsansätze	7
2.1. Kognitions-psychologische Überlegungen zur Urteilsbildung bei Befragungen	7
2.2. Methodische Möglichkeiten von Panel-Studien	9
2.3. Delphi-Technik als modifizierte Form der Gruppendiskussion	11
2.4. Expertenbefragungen	12
2.5. Anwendung von Elementen der Theorie rationalen Handelns auf die Delphi-Technik	13
2.6. Zusammenfassung	15
3. Untersuchungsprogramm	15
4. Hypothesen und Operationalisierung	16
5. Feldbericht	19
6. Auswertung	21
6.1. Allgemeines Konzept	21
6.2. Auswertung der Indikatorenschätzungen bei vierstufigen Skalen	21
6.3. Auswertung der Indikatorenschätzungen bei zweistufigen Antwortskalen	27
6.4. Auswertung der Mittelwertschätzungen bei Indikatoren mit siebenstufigen Antwortskalen	35
6.5. Auswertung der ergänzenden Expertenbefragung (vierte Welle)	43
7. Zusammenfassung und Ausblick	44
Literatur	47
Anlagenverzeichnis	50



## 1. Ausgangssituation

Die methodologischen Grundlagen der Delphi-Technik sowie deren konkrete methodische Umsetzung sind gegenwärtig mit zahlreichen Defiziten verbunden. So wurde z.B. bei einer Durchsicht der aktuellen Studienliteratur im Fach "Methoden der empirischen Sozialforschung" (Schnell/Hill/Esser 1992; Kromrey 1991; Atteslander 1984; v. Alemann 1984; Friedrichs 1984; Bortz 1984; Roth 1984; Hartmann 1979; Mayntz et al. 1978 und König 1973) nach Hinweisen auf die Delphi-Methode gesucht. Nur in zwei der zehn Bücher findet die Delphi-Methode Erwähnung: Bei Atteslander wird sie im Rahmen des Abschnittes "Leitfaden-Befragungen" auf 13 Zeilen abgehandelt, Bortz widmet ihr im Abschnitt "Schriftliche Befragung" eine Seite, auf der er allerdings für interessierte Leser auf den Sammelband von Linstone und Turoff "The Delphi Method - Techniques and Applications", die das Thema sehr ausführlich abhandeln und eine wesentliche Grundlage auch für die folgende Darstellung liefern, verweist. Damit muß zunächst vermerkt werden, daß es sich bei der Delphi-Methode offenbar um einen "Geheimtip" handelt, der in der deutschen allgemeineren sozialwissenschaftlich-methodischen Literatur nahezu unerwähnt bleibt.

Die Delphi-Methode hat ihren Ursprung in der amerikanischen Landesverteidigungsforschung. Sie wurde erstmals Anfang der 50er Jahre in einer "Project Delphi" genannten, von der Air Force gesponsorten Rand Corporation Studie eingesetzt. Diese sollte mögliche sowjetische Angriffsziele auf die USA sowie die dazu erforderlichen Atombomben ermitteln (vgl. Linstone/Turoff 1975, S. 10). Für die Lösung dieses Problem wurde die spezielle Methode der schriftlichen strukturierten Experten-Befragung (später "Delphi-Methode") entwickelt. Diese Entwicklung war insofern zwangsläufig, als eine Alternative zur Bearbeitung der für die Verteidigungsbereitschaft der USA äußerst brisanten Problemstellung mit konventionellen Methoden nicht existierte. Öffentlich bekannt geworden sind dann allerdings Delphi-Studien für lang- und mittelfristige Prognosen bei denen die Experten zunächst gebeten wurden, zukünftig denkbare Entwicklungen auf diesen Gebieten vorzuschlagen. Danach hatte die Expertengruppe in mehreren Wellen die Eintrittswahrscheinlichkeiten dieser Entwicklungen zu schätzen.

In Deutschland kommt Albach das Verdienst zu, mit seinem Artikel "Informationsgewinnung durch strukturierte Gruppenbefragung: Die Delphi-Methode" in der Zeitschrift für Betriebswirtschaft 1970 die Ergebnisse des Rand Corporation-Studie vorgestellt und damit die Methode einer breiteren Öffentlichkeit publik gemacht zu haben. Daß allerdings in der deutschen Soziologie nur punktuell von dieser Methode Notiz genommen wurde, konnte bereits in einem Literaturbericht (Häder/ Häder 1994) gezeigt werden.

In einer Reihe von Publikationen gibt es Versuche, die Delphi-Methode zu definieren. Dabei werden, je nach Profession der Autoren, naturgemäß jeweils bestimmte Sichtweisen in den Vordergrund gestellt. So taucht im Rahmen solcher Definitionen häufig die Kopplung von Delphi

und Prognosen auf (z.B. Overbury 1968; Rowlands 1969; Gewalt 1972; Saliger/Kunz 1981; G. Köhler 1992; Atteslander 1984; Bronner et al. 1991).

Beispielsweise sei hier die Begriffsbestimmung von G. Köhler zitiert:

"Die Delphi-Methode ist ein Verfahren, um aus Expertenmeinungen Prognosen zu gewinnen und Konsens und Dissens zwischen den Expertenmeinungen deutlich zu machen." (Köhler 1992, S.325)

Linstone/Turoff dagegen wählen bewußt eine sehr allgemeine Definition, die berücksichtigt, daß die Anwendbarkeit dieser Technik weit über Prognosezwecke hinausgeht:

"Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem." (Linstone/Turoff 1975, S.3)

Ähnlich ausgerichtet ist die Begriffsbestimmung von Bortz, die unserem Verständnis der Methode am besten entspricht:

"Es handelt sich hierbei um eine hochstrukturierte Gruppenkommunikation, deren Ziel es ist, aus den Einzelbeiträgen der an der Kommunikation beteiligten Personen Lösungen für komplexe Probleme zu erarbeiten." (Bortz 1984, S.189)

Bei aller Vielfalt der Anwendungsbereiche und der methodischen Abwandlungen besteht doch bei der weit überwiegenden Zahl der Delphi-Studien eine Gemeinsamkeit darin, daß der Zeithorizont der zu schätzenden Sachverhalte in der Zukunft liegt, d.h. es sich mehrheitlich um Prognose-Delphis handelt. Nach unserer Schätzung betrifft dies weit über 90 Prozent der uns bekannten Anwendungsfälle, in Deutschland gar 100 Prozent.

Die Delphi-Methode findet, so zeigen verschiedene Beispiele, also immer insbesondere dann Anwendung, wenn es um die Schätzung von Sachverhalten geht, die nicht direkt abgebildet werden können, da sie nicht aktuell präsent bzw. real existent sind (z.B. weil sie in der Zukunft oder Vergangenheit liegen) und von daher kreative Leistungen der Experten gefragt sind.

Die Delphi-Technik bedient sich dazu des folgenden Vorgehens:

1. Operationalisierung der allgemeinen Fragestellung mit dem Ziel, konkrete Kriterien für eine spätere Beurteilung zu gewinnen. Dieser erste Schritt kann sowohl von einem Forscherteam als auch mit Hilfe einer offenen Befragung von externen Experten bewältigt werden.
2. Ausarbeitung eines standardisierten Frageprogramms. Dieses dient dazu, Experten anonym nach ihren Meinungen zum interessierenden Sachverhalt zu befragen.
3. Aufbereitung der Befragungsergebnisse durch das Forscherteam und anonymisierte Rückmeldung der Ergebnisse an die beteiligten Befragten.
4. Wiederholung der Befragung auf der Grundlage der über die Rückinformationen gewonnenen Erkenntnisse bis zum Erreichen eines bestimmten Abbruchkriteriums.

Die kontrovers geführte Diskussionen um die Leistungsfähigkeit der Delphi-Methode wird im folgenden Zitat klar: "Die Einstellung von Literatur und Praxis zur Delphi-Methode ist zwiespältig und läßt sich treffend durch ein These-Antithese-Verhältnis charakterisieren. Von ihren Begründern und deren zahlreichen Epigonen wird sie überwiegend unkritisch dargestellt und teilweise ohne Berücksichtigung der jeweiligen situativen Bedingungen relativ stereotyp eingesetzt. Lediglich aufgrund ihres Potentials, die in face-to-face Gruppen möglichen negativen sozialpsychologischen Faktoren zu vermeiden, propagiert beispielsweise Schöllhammer 'die Delphi-Methode als optimale Form der Gruppenprognose'. Sackmans teilweise polemisch geführte 'Delphi Critique' kann als d i e Antithese interpretiert werden: 'It is time for the oracles to move out and for science to move in'. Dieses These-Antithese-Verhältnis bedarf einer systematischen und, soweit möglich, auch empirisch fundierten Synthese. (Wechsler 1978 S. 596) An anderer Stelle wird diese Diskussion ausführlicher abgehandelt (vgl. Häder/ Häder 1994).

Generell gilt es folgendes zu bemerken:

In Deutschland spielt Delphi bisher im Vergleich zu anderen sozialwissenschaftlichen Methoden nur eine untergeordnete Rolle. Es existieren fast keine Beiträge, die Anwendungen von Delphi bzw. Beiträge zur Methodenentwicklung zum Inhalt haben. Eine zusammenfassende Darstellung der Technik in deutscher Sprache (vergleichbar etwa mit dem Band von Linstone/Turoff 1975) liegt ebenfalls nicht vor.

Diese Stellung der Methode - wie sie vermittelt über ihren Platz in der Literatur deutlich wird - ist unseres Erachtens der potentiellen Leistungsfähigkeit von Delphi auch in der Soziologie keineswegs adäquat, dies zeigen die vielfältigen erfolgreichen Anwendungen und Experimente z.B. in den USA und in der deutschen Betriebswirtschaft.

Um jedoch einen anerkannten Platz unter den Methoden der empirischen Sozialwissenschaft zu erlangen, bedarf die Delphi-Technik allerdings einer ausgiebigeren Testung. Damit ein größeres Vertrauen in diese Methode aufgebaut werden kann, sind ihre Vorzüge und Grenzen unter soziologischem Anspruch darzustellen. Außerdem würde eine stärkere Publizität (angefangen bei der Ausildung von Soziologie-Studenten) diesem Instrument dienlich sein.

Damit verlagert sich die Problematik des breiteren Einsatzes von Delphi auf die Erklärung, für welche Art der Fragestellung welches Vorgehen geeignet ist. Es geht also nicht um eine grundsätzliche Legitimation von Delphi-Studien, sondern vielmehr darum, mit welcher konkreten Delphi-Strategie welches Problem gelöst werden kann.

Zusammenfassend läßt sich der Stand der methodischen Erkenntnisse zur Delphi-Technik folgendermaßen beschreiben:

- Die erkenntnistheoretischen Grundlagen für die Anwendung der Delphi-Technik sind, wahrscheinlich aufgrund des bislang vorrangigen Einsatzes von Delphi-Studien für



Langzeitprognosen auf den verschiedensten Gebieten, unverbindlich-vielfältig, nur wenig strukturiert und damit unzureichend ausgearbeitet.

- Ein systematischer Literaturüberblick zeigte, daß die Leistungsfähigkeit von Delphi-Studien für sozialwissenschaftliche Untersuchungsziele - besonders im deutschsprachigen Raum - methodisch nicht erforscht ist. Besonders Ansätze zur Validierung der Ergebnisse von Delphi-Studien sind - nicht zuletzt aufgrund des bisher vorherrschenden Anwendungsgebiets Langzeitprognosen - fast völlig unbekannt.
- Die zahlreichen erhebungsmethodischen Variationsmöglichkeiten bei der Nutzung der Technik sind weder durch methodische Tests abgesichert noch überhaupt ausreichend erschöpfend systematisiert.

Diese Bestandsaufnahme legitimiert ein Programm an methodischen Testerhebungen und methodologisch-vergleichenden Studien zur angemessenen Rezeption der Delphi-Technik in das sozialwissenschaftliche Methodenarsenal.

Bevor wir die Ergebnisse eines Tests vorstellen, der die Schätzung von Antworten bei einer Bevölkerungsbefragung (Shell Jugendstudie von 1992) durch ZUMA-Mitarbeiter beinhaltete, sollen zunächst theoretische Erklärungsansätze für die bei Delphi ablaufenden Prozesse dargestellt werden.

## **2. Erklärungsansätze**

Die Delphi-Technik knüpft methodologisch an verschiedene Ansätze an. Dies kann bei der notwendigen Suche nach Validierungsmöglichkeiten von Bedeutung sein. Zu nennen sind Arbeiten zu:

1. kognitions-psychologischen Fragen,
2. der Methode Gruppendiskussion und zur Gruppendynamik,
3. Panel-Studien sowie
4. Expertenbefragungen.

Diese bilden den Hintergrund für die methodischen Tests zur Delphi-Technik. Auf einige Aspekte soll zunächst vertiefend eingegangen werden.

### **2.1. Kognitions-psychologische Überlegungen zur Urteilsbildung bei Befragungen**

Methodologisch gilt es zunächst zu begründen, warum Panel-Befragungen mit Experten überhaupt zu einer Verbesserung der ursprünglich abgegebenen Urteile führen können, wenn von Welle zu Welle den Experten die erzielten Ergebnisse rückgemeldet werden. Dazu soll der

Prozeß der Urteilsbildung bei einer Befragung aus kognitions-psychologischer Sicht näher betrachtet werden:

"Die meisten Forscher stimmen darin überein, daß Befragte mehrere Aufgaben lösen müssen: Sie müssen zunächst die Frage verstehen, relevante Information aus dem Gedächtnis abrufen und auf der Basis dieser Information ein Urteil bilden. Schließlich müssen sie ihr Urteil in die vorgegebenen Antwortalternativen einpassen und es unter Umständen 'editieren', um Erwägungen sozialer Erwünschtheit Rechnung zu tragen." (Schwarz 1991, S. 71)

Handelt es sich wie in unserem - später noch explizit darzustellenden - Test um Meinungsfragen, "müssen sie ein früher gebildetes Urteil erinnern oder relevante Information aus dem Gedächtnis abrufen, die ihnen die Bildung eines Urteils zum Befragungszeitpunkt erlaubt." (Schwarz et al. 1991, S.176)

Hier bilden Arbeiten zur flexiblen Konstruktion mentaler Repräsentationen die theoretische Grundlage (vgl. Barsalou 1987, 1989). Bei der Abgabe eines evaluierenden Urteils wird auf das entsprechende Wissen zurückgegriffen. Wichtig für Delphi-Studien ist zunächst, daß dem Befragten verschiedene Arten von Informationen zur Verfügung stehen. Folgendes längere Zitat stellt u.E. eine gute Zusammenfassung kognitions-psychologischer Gedanken dar:

"Einige Informationen werden ... dabei wahrscheinlich unter allen Umständen in den Sinn kommen ... Diese Informationen sind im Gedächtnis *chronisch verfügbar* und ihr Abruf ist daher *kontextunabhängig* (vgl. Higgins/King 1981). Andere Informationen werden ... nur unter bestimmten Umständen einfallen. ... Solche Informationen sind nur *temporär verfügbar*. ... Wir nennen diese Information deshalb '*kontextabhängig*'- sie kommt Befragten nur in einem bestimmten Kontext in den Sinn. Dies spiegelt wider, daß wir nie alle potentiell relevante Information aus dem Gedächtnis abrufen, sondern den Suchprozeß abbrechen, sobald wir genügend Information erinnert haben, um mit hinreichender Sicherheit ein Urteil bilden zu können (vgl. Bodenhausen/ Wyer 1987). Das Urteil beruht daher nur auf der Teilmenge potentiell relevanter Information, die zum Urteilszeitpunkt leicht verfügbar ist - sei dies chronisch oder temporär. Während kontextunabhängige, chronisch verfügbare Information für eine gewisse *Stabilität* im Urteil sorgt, bewirkt temporär verfügbare, kontextabhängige Information eine gewisse *Variabilität* im Urteil." (Schwarz 1991, S. 72)

Als Experten im Sinne der Delphi-Technik können danach jene Personen bezeichnet werden, die zunächst über eine besondere Qualität an chronischen Informationen verfügen. Die den Experten dann zusätzlich noch zurückgemeldete Information soll bewirken, daß diese den "Suchprozeß" in ihrem Gedächtnis nochmals aufnehmen und nach weiteren für den jeweiligen Sachverhalt relevanten Informationen fahnden. Kognitions-psychologisch bewirkt die rückgemeldete Gruppenantwort einen gewissen Kontexteffekt. Hierzu liegen eine ganze Reihe an Experimenten vor (vgl. z.B. Schwarz 1991).

Bestimmte Tests ergaben beispielsweise Hinweise auf die informative Funktion von Antwortvorgaben. Auch diese sind für unsere Delphi-Problematik von Bedeutung. Danach nehmen Befragte an, "der Wertebereich der Antwortvorgaben gäbe das Wissen des Forschers

über die Verteilung des erfragten Verhaltens wieder. Gemäß dieser Annahme wird die durchschnittliche oder 'typische' Verhaltenshäufigkeit durch Werte im Mittelbereich der Skala abgebildet, während die Extreme der Skala den Extremen der Verteilung entsprechen." (Schwarz et al. 1991, S. 177) Hieraus müßte man vermuten, daß analog bei Delphi-Studien, in denen die Schätzung von Antwortverteilungen von den Experten abgefordert wird, vor allem Normalverteilungen richtig geschätzt werden können. Diesem Hinweis wird später noch weiter nachgegangen werden.

Diese kognitions-psychologischen Überlegungen zusammenfassend gilt es festzustellen: Die Befragten verfügen über mehr Wissen, als sie (zunächst) für die Beantwortung einer Frage benutzen. Dieses ursprünglich nichtaktivierte Rest-Wissen soll mit Hilfe der Delphi-Technik genutzt werden. Dazu sind die geeigneten methodischen Verfahren zu erarbeiten.

Zwei weitere Aspekte sind interessant:

Strack (1991, S.18ff.) berichtet über Experimente, bei denen Fragen nach fiktiven Sachverhalten gestellt wurden und die (trotzdem) von den Probanden beantwortet wurden, z.B. um Unwissen zu verbergen. Interessant ist in unserem Zusammenhang an diesen Experimenten, daß die Antworten in solchen Fällen nicht zufallsgesteuert erfolgen, sondern sich z.B. aus dem Kontext der übrigen Befragung erklären lassen. Einen "geistigen Münzwurf" mit dessen Hilfe "Zufallsantworten" erzeugt werden gibt es demnach nicht. Auf die Antworten der Experten bei Delphi-Studien angewandt würde dies bedeuten, daß auch im Falle eines "non-attitudes" von den Experten Vorgaben entsprechend eines von ihnen angenommenen Kontextes geschätzt werden. Außerdem könnte angenommen werden, daß selbst bei komplizierten Aufgaben mit geringen Ausfallquoten gerechnet werden kann.

Eine Fehlerquelle bei Befragungen ist die Motivation des Probanden, sozial erwünschte Antworten geben zu wollen. Dies führt dazu, daß Befragte, die ein starkes Bedürfnis verspüren, sozial erwünscht zu antworten, lügen, um einen besseren Eindruck zu hinterlassen (vgl. Strack 1991, S.17). Dieser Einfluß auf das Befragtenverhalten wird jedoch im Rahmen unseres Tests wahrscheinlich nicht auftreten, da nicht die eigene Meinung des Experten, sondern die einer anderen Population erfragt wird, für die der Experte nicht "verantwortlich" ist. Sollte sozial erwünschtes Antworten in unserem Test überhaupt eine Rolle spielen, so am ehesten als Wunsch gegenüber den Veranstaltern, die Ergebnisse als Experten besonders gut schätzen zu wollen.

## **2.2. Methodische Möglichkeiten von Panel-Studien**

Friedrichs (1980, S. 368ff.) nennt eine Reihe von Vorteilen, die gerade Panel-Studien auszeichnen. Vor allem die folgenden sind auch für die Beurteilung der Delphi-Technik relevant: Solche Vorgehensweisen ermöglichen die Analyse auf der Ebene der Personen, die ihre Einstellungen und/oder ihr Verhalten ändern; die "Analyse der Ursachen (Bedingungen) des

Wandels in einzelnen Bereichen auf die Gleichzeitigkeit/Ungleichzeitigkeit des Wandels in verschiedenen Bereichen, auf die Interdependenz von Prozessen des Wandels" (S. 368).

Bei Bortz (1984) finden sich weitere Vorzüge dieser Technik: "Die anfängliche Unsicherheit, sich in einer ungewohnten Befragungssituation zu befinden, verliert sich im Verlaufe der Zeit, der Befragte lernt, seine eigenen Ansichten und Meinungen treffsicherer und genauer zu formulieren; die anfängliche Skepsis, persönliche Angaben könnten trotz zugesicherter Anonymität mißbräuchlich verwendet werden, schwindet, das Panelmitglied entwickelt so etwas wie ein Verantwortungsbewußtsein dafür, daß die zu treffenden Entscheidungen auf korrekten Planungsunterlagen beruhen etc." (S.314)

Nehnevajsa (1967, S. 202ff.) verweist außerdem auf die bei Panel-Studien für eine detailliertere Auswertung erforderlichen Zusatzinformationen über die "fluktuierenden" Einheiten. (Dieser Anregung folgen wir besonders mit der abschließenden Erhebung über die von den Experten bei der Studie gewonnenen Erfahrungen - siehe Abschnitt 6.5.) Für die Evaluation der Delphi-Technik bedeutet dies, zusätzlich Informationen zu erheben, mit denen beispielsweise die möglichen Determinanten für eine (bessere oder schlechtere) Schätzung der Antwortverteilungen von Befragungen ermittelt werden können. Es geht letztlich darum, die Voraussetzungen zu erforschen, unter denen Experten bei Delphi-Studien ihre Urteile verbessern.

Nehnevajsa unterscheidet bei Panel-Studien konstante, intermittierende und kovariierende Einflußfaktoren, die die Fluktuation der Antworten erklären können. Konstante Einflußfaktoren sind solche, "die sich für die Dauer der Panel-Untersuchung nicht ändern." (S.203) Diese dürften konkret in unserem Zusammenhang die entscheidenden sein. Dabei könnte es sich beispielsweise um die Fachrichtung des jeweiligen Experten handeln. Intermittierende Einflußfaktoren sind "Ereignisse, die zwischen den beiden (oder zwischen zwei) Panel-Befragungen stattfinden." (S.204) Dabei könnte es sich um eine Erweiterung des Wissens z.B. durch zwischen den Wellen gelesene Literatur o.ä. handeln. Schließlich werden mit kovariierenden Einflußfaktoren jene bezeichnet, "die sich im Laufe der Untersuchung selbst ändern." (S.205) Wenn die einzelnen Wellen der Delphi-Befragung zeitlich dicht beieinander liegen, gehen wir davon aus, daß sowohl intermittierende als auch kovariierende Einflußfaktoren - mit Ausnahme der ausdrücklich beabsichtigten Wirkung der rückgemeldeten Ergebnisse natürlich - zu vernachlässigen sind.

Weiter werden analog dazu a) bedingte Fluktuationen analysiert (z.B. Fluktuationstabellen für erfolgreich und weniger erfolgreich schätzende Experten getrennt aufgestellt), b) bedingte Wirkungen festgestellt (hier lautet die Fragestellung: "Wie sind Befragte, die irgendeinem Ereignis ausgesetzt sind, zu vergleichen mit anderen Beteiligten, die von diesem Ereignis nicht beeinflußt wurden?" (S.209)) und schließlich c) die Analyse gegenseitiger Wirkungen durchgeführt, bei der die kovariierenden Einflußmerkmale betrachtet werden. Bei Panel-

Befragungen mit mehr als zwei Wellen können zusätzlich kontinuierlicher Wandel und Oszillation untersucht werden.

Aber auch für die Delphi-Technik relevante Nachteile und Grenzen von Panel-Studien dürfen nicht übersehen werden: die Mortalität der Stichprobe und die fehlende Anonymität der Befragten zwar nicht untereinander, aber gegenüber dem Monitorteam. Ein anderer, in der Methodenliteratur aufgeführter Nachteil von Panel-Studien ist die Beeinflussung der Teilnehmer durch die Methode (vgl. z.B. Bortz 1984, S.188). Dieser Aspekt ist jedoch bei Delphi ausdrücklich angestrebt und entfällt damit als Problem. Schließlich ist ein weiterer Effekt nicht auszuschließen: "Das Panelmitglied fühlt sich maßgeblich am Zustandekommen der Befragungsergebnisse beteiligt und kann deshalb danach trachten, das Resultat der Befragung in eine ihm erstrebenswert erscheinende Richtung zu lenken, indem es extreme, von der Realität abweichende Positionen vertritt." (Bortz 1984, S. 314)

Zusammenfassend ist festzustellen, daß die Tatsache, daß sich der Delphi-Ansatz einer Panel-Befragung bedient besondere Bedeutung für die Darstellung des Meinungsbildungsprozesses der Experten hat und damit für die Evaluation dieser Methode von Wert ist.

### **2.3. Delphi-Technik als modifizierte Form der Gruppendiskussion**

Die Delphi-Technik erfüllt einige Kriterien der in ihren Vor- und Nachteilen methodisch relativ gut erforschten Gruppendiskussionen.

Für die Gruppendiskussion im allgemeinen spricht u.a., "daß die Gruppengesprächssituation eher als die des Einzelinterviews zur Aktualisierung und Explikation 'tieferliegender' Bewußtseinsinhalte stimuliere." (Mangold 1967, S. 230; vgl. auch Friedrichs 1980, S. 246)

Gegen die Gruppendiskussion wird angeführt, "daß bei der Anwendung von Gruppendiskussionen in der Regel bedeutend größere Ausfälle auftreten als in der mündlichen Umfrage" (Mangold 1967, S.232; vgl. auch Abrams 1949, Smith 1955, Pollock 1955).

"Als zweite grundsätzliche Schwierigkeit wird in der vorliegenden Literatur die ungleichmäßige Beteiligung der einzelnen Versuchspersonen an den Diskussionen hervorgehoben." (Mangold 1967, S.232)

Beide Einwände gegen Gruppendiskussionen treten bei deren Modifikation als Delphi-Technik nicht auf, während der zuvor angeführte Vorteil gleichzeitig erhalten bleibt.

Experimente zu sozialpsychologischen Prozessen bei Gruppenbefragungen, die auch für die Delphi-Technik relevant sind, betrafen verschiedene Aspekte: Einige "zeigen, daß informelle Gruppenmeinungen als Maßstäbe für die eigene Meinungsbildung dienen, wenn andere Maßstäbe nicht zur Verfügung stehen oder unglaubwürdig werden, ... (vgl. Muzafer Sherif 1937) ... Er konnte experimentell zeigen, daß mangels anderer Maßstäbe die Urteile der einzelnen

Versuchspersonen vollends auf die Meinungen der anderen Versuchspersonen angewiesen waren." (Mangold 1967, S.237)

Auch dieser Hinweis ist für die Delphi-Technik wichtig. Er gibt Anlaß zu der Vermutung, daß diejenigen Experten, deren Meinungsbildung abgeschlossen ist, sich nicht an der Gruppenmeinung orientieren, während die genannten unentschlossenen Versuchspersonen sich an denen orientieren, die über vermeintlich gesichertes Wissen verfügen. Unsere Experimente werden zeigen, inwieweit es diesen Versuchspersonen dann auch gelingt, gute Schätzungen abzugeben.

Ein anderes Beispiel ist ein amerikanisches Experiment zu Beginn der 60er Jahre. "So kamen Warren Breed und Thomas Ktsanes (1961) in einer Untersuchung in New Orleans zu dem Ergebnis, daß die Zahl der Anhänger einer Beibehaltung der Segregation von Weißen und Negeren weit überschätzt wurde, und zwar am stärksten von denen, die die Segregation befürworten." (Mangold 1967, S.239) Hier wird auf eine mögliche auch für die Delphi-Technik relevante Fehlerquelle verwiesen. So besteht offenbar die Tendenz, die eigene Meinung zu überschätzen, auch wenn es z.B. um die Beurteilung "äußerer" Sachverhalte geht, bei der die eigene Ansicht (eigentlich) ausgeschlossen sein sollte.

## **2.4. Expertenbefragungen**

Bei Delphi-Studien kommt es zunächst darauf an, eine Expertengruppe zusammenzustellen, die dazu in der Lage sein soll, ein besonders sachkundiges Urteil abzugeben. Die Befragung von Experten wird in der Literatur mitunter als eine (Sonder-)Form der Befragung geführt (vgl. Atteslander 1984, S. 119). Besondere methodische Hinweise sind jedoch für die Handhabung dieser speziellen Erhebungstechnik nicht erforderlich. Wichtig erscheint lediglich der Verweis darauf, daß die Befragung von Experten ein sozialwissenschaftlich anerkanntes Verfahren mit eigener Leistungsfähigkeit ist. Der jeweilige Gegenstand bestimmt auch hier die Ausgestaltung des Vorgehens.

Bereits eine gewisse Tradition besitzt die Konsultation von Experten im Rahmen der Skalierungsverfahren. So legten Thurstone und Chave (1929) einer Expertengruppe verschiedene Items zur Beurteilung der jeweils in ihnen zum Ausdruck kommenden Merkmalsausprägung vor. Erfahrungen belegen, daß bei der Erarbeitung von Fragebogen für Expertenbefragungen einige Probleme (z.B. Vermeidung von Fachbegriffen) entfallen.



## 2.5. Anwendung von Elementen der Theorie rationalen Handelns auf die Delphi-Technik

Die Theorie des rationalen Handelns findet verstärkt Anwendung bei der Erklärung sozialer Prozesse. In einem Aufsatz wendet z.B. Esser (1986) diese Theorie auch auf Situationseinflüsse beim Interview an. Auch dieser Ansatz kann interessante Ansatzpunkte für eine Systematisierung von Determinanten des Befragtenverhaltens liefern. Es soll deshalb an dieser Stelle mit Hilfe der Theorie rationalen Handelns versucht werden, einige Verhaltenselemente von Experten bei einer Delphi-Befragung zu ergründen. Im Mittelpunkt steht dabei eine Systematisierung von Einflußgrößen und nicht der Versuch, mit dieser empirisch bisher kaum umgesetzten Theorie etwa den Erfolg einer Schätzung zu bestimmen. Analog zu Esser (1986, S. 322) unterscheiden wir dazu: von den Probanden jeweils unterschiedlich bewertete Ziel-Situationen ( $U_j$ ), einen gewissen Satz von ihnen zur Verfügung stehenden Handlungsalternativen ( $A_j$ ) sowie bestimmte subjektive Erwartungen ( $p_{ij}$ ), mit den einzelnen Handlungen das jeweilige Ziel zu erreichen. Konkret nehmen wir an, daß für die Experten (es soll an dieser Stelle nur die für die Delphi-Technik charakteristische zweite und dritte Welle interessieren) zwei Handlungsalternativen bestehen:

$A_1$ : Beibehaltung der in der vorangegangenen Welle geäußerten Schätzung oder

$A_2$ : Korrektur der ursprünglichen Meinung.

Mit dieser Beschränkung auf eine Handlungsalternative halten wir das Modell bewußt auf einem relativ einfachen Niveau. Zugleich besteht sukzessive die Möglichkeit, die Experten eindeutig empirisch  $A_1$  oder  $A_2$  zuzuordnen.

Wir nehmen - ebenfalls weiter stark vereinfacht - an, daß es für die Experten lediglich zwei Zielsituationen gibt:

$U_1$ : Anpassung an die explizit rückgemeldete "Gruppennorm" oder

$U_2$ : Orientierung am eigenen Wissen bzw. an den eigenen Erfahrungen in Bezug auf den erfragten Sachverhalt relativ unabhängig vom Gruppenergebnis.

Es wird damit z.B. vorausgesetzt, daß eine prinzipielle Kooperationsbereitschaft besteht und sie damit nicht als Ziel in dieses Modell aufgenommen werden muß. Ausgeschlossen bleiben außerdem solche Ziele wie Bestreben nach sozialer Anerkennung oder bewußte Manipulation der Ergebnisse in eine bestimmte Richtung. Auch die beiden zuerst genannten Handlungsziele lassen sich empirisch nachvollziehen.

Nun bilden - der Theorie rationalen Handelns entsprechend - die Experten bestimmte Werte für die Wahrscheinlichkeiten, mit den einzelnen Handlungen die einzelnen Ziele zu realisieren.

Konkret handelt es sich um:

$p_{11}$ : der Gruppennorm dadurch gerecht zu werden, daß man seine ursprüngliche Antwort beibehält;

$p_{12}$ : das eigene Wissen dadurch darzustellen, daß man seine ursprüngliche Antwort beibehält;

$p_{21}$ : der Gruppennorm dadurch gerecht zu werden, daß man seine ursprüngliche Antwort ändert;

$p_{22}$ : das eigene (eventuell inzwischen geänderte) Wissen dadurch darzustellen, daß man seine Antwort verändert.

"Es wird nun weiter angenommen, daß der Akteur für jede einzelne Handlungsalternative  $A_j$  in bezug auf jedes Ziel eine Gewichtung mit der zugehörigen subjektiven Wahrscheinlichkeit vornimmt." (Esser 1986, S. 322) Auf diese Weise wird eine individuelle Nutzenerwartung für jede Handlung ermittelt. Dazu werden  $U_j$  und  $p_{ij}$  multipliziert und danach die entstandenen Werte summiert (Summe  $p_{ij}U_j$  über alle  $j$  wird gebildet). Sollte beispielsweise eine bestimmte Erwartung ( $p_{ij}$ ) oder eine bestimmte Bewertung ( $U_j$ ) gering (oder entsprechend hoch) sein, so ist auch der Nutzen einer Handlung gering (bzw. hoch). Schließlich wird jene Handlung gewählt, die für die Experten die höchste Nutzenerwartung beinhaltet.

Beispielsweise ist der Fall denkbar, daß ein Experte meint, der Gruppennorm dadurch besonders gut gerecht werden zu können, daß er seine Schätzung stärker an das Gruppenergebnis anpaßt - z.B. weil seine ursprüngliche Schätzung außerhalb des rückgemeldeten Streuungsbereichs liegt. Der Wert für  $p_{21}$  wäre in diesem Fall hoch. Zugleich soll angenommen werden, daß dieser Experte meint, sich nicht an der Gruppennorm orientieren zu müssen, z.B. weil er sich mehr als andere inhaltlich mit der Thematik befaßt hat;  $U_1$  wäre in diesem Fall gering. Dann ist auch das gebildete Produkt gering, d.h. der Nutzen, der für den Experten aus einer Meinungsänderung resultieren würde, wäre in diesem Fall nicht sonderlich hoch.

Um konkretere Erklärungen für das Zustandekommen von  $A_1$  oder  $A_2$  abgeben zu können bzw. um die Determinanten des Antwortverhaltens mit dieser Theorie zu bestimmen, müssen nun konkrete Annahmen über die übrigen Variablen getroffen werden. Zuvor soll nochmals betont werden: Es geht hier nicht um die Vorhersage eines bestimmten Antwortverhaltens mit Hilfe der Theorie rationalen Handelns, sondern um eine Systematisierung der Determinanten des Antwortverhaltens.

- Zunächst erscheint es plausibel, daß die Größen von  $p_{11}$  und  $p_{21}$  von der Rückmeldung des Gruppenergebnisses abhängen. Der jeweilige  $p$ -Wert ist also dann besonders groß, wenn die eigene ursprünglich geäußerte Meinung der Gruppennorm entspricht ( $p_{11}$ ) bzw. wenn sie außerhalb des rückgemeldeten Streuungsbereiches liegt ( $p_{21}$ ).
- Weiterhin ist leicht feststellbar, ob die Experten  $A_1$  oder  $A_2$  gewählt haben.
- Unbekannt bei der Determination des Antwortverhaltens sind nun zunächst noch die Intensitäten der Bewertung der Zielsituationen ( $U_j$ ). Diese lassen sich ermitteln mit Hilfe der Zusatzbefragung nach Abschluß aller eigentlichen Delphi-Wellen (vgl. Abschnitt 6.5. des vorliegenden Berichts). Als Indikator für  $U_1$  (Anpassung an die explizit rückgemeldete Gruppennorm) kann die Frage benutzt werden: "Ich habe über den Abstand meiner Schätzung vom Gruppenmittel nachgedacht."  $U_1$  würde damit einen Wert zwischen 1 und 7 annehmen.
- $U_2$  (Orientierung am eigenen Wissen bzw. an den eigenen Erfahrungen) kann ermittelt werden mit Hilfe der Frage: "Ich habe (nochmals) über den Inhalt nachgedacht." Auch hier würde wieder eine von 1 bis 7 reichende Skala vorgegeben.



- Schließlich sind die Wahrscheinlichkeiten (p<sub>12</sub> und p<sub>22</sub>), das eigene Wissen mit einer konstanten oder veränderten Antwort darzustellen, empirisch zu ermitteln.

Im Zusatzbogen wurde gefragt: "Ich war bemüht, bei den verschiedenen Wellen möglichst ein konstantes Urteil abzugeben." Diese Frage soll als Indikator sowohl für p<sub>12</sub> (Zustimmung) als auch als Hinweis auf p<sub>22</sub> (Ablehnung) gelten.

Damit wird das Gleichungssystem prinzipiell lösbar. Es besteht so die Möglichkeit, den Meinungsbildungsprozeß der Experten bei der Delphi-Methode transparenter zu machen. Versuche, diese Überlegungen empirisch zu belegen, würden allerdings den Rahmen dieser Darstellung erster Ergebnisse sprengen und sollen deshalb späteren Arbeiten vorbehalten bleiben.

## **2.6. Zusammenfassung**

Die sozialwissenschaftliche Methodologie enthält - dies wurde bereits an anderer Stelle dargestellt - keine geschlossene Darstellung der methodischen Prinzipien für die Handhabung der Delphi-Technik (vgl. Häder/ Häder 1994). Genutzt werden können dagegen Elemente der hier genannten Techniken, mit denen die Delphi-Methode "verwandt" ist. Dabei konnten eine ganze Reihe von Anhaltspunkten für das Funktionieren der Delphi-Technik zusammengestellt werden. Insbesondere für die empirische Validierung von Delphi-Ergebnissen kann auf solche Erkenntnisse zurückgegriffen werden.

## **3. Untersuchungsprogramm**

Aus dem eingangs nur kurz umrissenen umfangreichen Problemkatalog lassen sich - jeweils in einzelnen Testerhebungen - stufenweise (nur) ausgewählte Probleme bearbeiten. Die methodischen Tests zur Delphi-Technik erfordern damit zahlreiche, jeweils aufeinander abgestimmte empirische Untersuchungen. Diese können an die genannten theoretischen Ansätze anknüpfen.

Als erster Schritt und damit als Einstieg ist im Frühjahr 1994 eine Testbefragung mit ZUMA-Mitarbeitern veranstaltet worden. Diese wird im vorliegenden Bericht ausgewertet. Zwei Problemkomplexe sind im Rahmen des ersten Tests bearbeitet worden:

### **1. Validierung von Ergebnissen einer Delphi-Studie.**

Die häufige Nutzung von Delphi-Studien für Langzeitprognosen machte Validierungsansätze bisher nahezu unmöglich. Tests, die belegen, daß die Delphi-Technik zu validen Ergebnissen führt, sind jedoch unerläßlich, um Vertrauen in die Methode zu erzeugen bzw. um sie überhaupt wissenschaftlich zu legitimieren. Validierungsbemühungen müssen damit zunächst im Mittelpunkt der methodischen Tests stehen. Gleichzeitig soll dabei gezeigt werden, daß die Delphi-Technik

originäre Vorzüge gegenüber anderen Verfahren (z.B. einmalige Expertenbefragung oder face-to-face Gruppendiskussion) besitzt.

Eine Validierungsmöglichkeit bietet die Nutzung der Delphi-Technik für die Schätzung von überprüfbaren - jedoch weitgehend unbekannten - Sachverhalten, wie beispielsweise die Ergebnisse einer Bevölkerungsbefragung. Mit der Schätzung numerischer Angaben (z.B. Mittelwerte und Antwortverteilungen) wird ein für bisherige Delphi-Studien typisches Problem aufgegriffen, so daß von daher die in unserem Experiment gewonnenen Ergebnisse auch verallgemeinert werden können.

## **2. Die Prüfung der Anwendbarkeit der Delphi-Technik für die Schätzung von Ergebnissen repräsentativer Bevölkerungsbefragungen.**

Die tatsächliche Breite der inhaltlichen Anwendungsmöglichkeiten für Delphi-Studien ist derzeit unklar. Auf eine ganze Reihe thematisch mitunter völlig unterschiedlicher Delphi-Projekte ist bereits an anderer Stelle (Häder/ Häder 1994) verwiesen worden. Zumindest die vorliegenden Ansätze verfolgen also sehr heterogene Ziele. Der Nachweis, unter welchen Bedingungen sich auch die Ergebnisse von Bevölkerungsbefragungen mit der Delphi-Technik schätzen lassen, würde eine weitere originelle Anwendungsrichtung begründen. Diese hätte praktischen Nutzen insbesondere für retrospektive Studien bzw. zur Ermittlung von Sachverhalten, über die ausreichende Daten nicht vorliegen oder - warum auch immer - empirische Bevölkerungsbefragungen als Methode nicht angewandt werden können.

Für die Bearbeitung dieser beiden - in eine gemeinsame Fragestellung integrierten - Problemkomplexe stellte eine Befragung am ZUMA eine Lösungsmöglichkeit dar. Am ZUMA tätige Mitarbeiter besitzen aufgrund ihres Tätigkeitsprofils ausreichend Kompetenz, um über die Ergebnisse von Befragungen in einer Delphi-Studie Auskunft zu geben.

Um möglichst viele methodisch relevante Aussagen gewinnen zu können, wurden verschiedene Aufgabentypen eingesetzt: die Schätzung vierstufiger Verteilungen, die Schätzung von Antwortverteilungen bei Alternativfragen sowie die Schätzung von Mittelwerten aus siebenstufigen Antwortvorgaben. Als inhaltlich für die ZUMA-Experten vermutlich ausreichend nachvollziehbar wurden Indikatoren zum Zeitbewußtsein (fünf Fragen mit vierstufiger Verteilung sowie vier Alternativfragen) und Wichtigkeitsurteile (Mittelwerte bei 18 siebenstufigen Antwortvorgaben) aus einer Erhebung von 1991 bei 13- bis 29-jährigen Jugendlichen verwandt.

## **4. Hypothesen und Operationalisierung**

Die Expertenschätzungen von Ergebnissen einer Bevölkerungsbefragung lassen sich mit Hilfe des Vergleichs mit den tatsächlichen Ergebnissen dieser Befragung eindeutig verifizieren bzw. falsifizieren. Es ist möglich, über die verschiedenen Wellen hinweg die Entwicklung der Schätzungen zu verfolgen. Dazu bieten sich zwei Kriterien: Es interessiert (1.) die abschließende,

insgesamt erreichte Treffgenauigkeit sowie (2.) die Annäherung (oder Entfernung) der Schätzungen an den richtigen Wert zwischen den verschiedenen Wellen. Eine gewisse Streuung der Abweichungen zwischen den Experten jeweils vorausgesetzt, wird damit die Bearbeitung von Hypothesen über die Einflußgrößen für "gute Schätzungen" - oder gute Delphi-Schätzexperten - bzw. für "erfolgreiche Annäherungen an den richtigen Wert" möglich.<sup>1)</sup>

Drei Grundhypothesen sind prinzipiell ableitbar:

1. Der Erfolg der Delphi-Schätzungen ist abhängig von den beteiligten **Experten** (Expertenhypothese), von deren Erfahrungen, von deren sozialer Integration, ihrem Abstraktionsvermögen usw.
2. Die Güte der Schätzungen resultiert aus der Spezifik der zu schätzenden **Sachverhalte**. (Inhaltshypothese)
3. Die Annäherungen der Schätzungen an den tatsächlichen Wert bzw. die Treffgenauigkeit resultieren darüber hinaus aus der konkreten Gestaltung der **Methode**. (Methodenhypothese)

Bei dem in der ersten Teststufe benutzten Design wird eine über alle Wellen konstante Methode eingesetzt, auf einen Split der Methode ist verzichtet worden. Damit können solche Fragen wie die Abhängigkeit der Schätzungen etwa von der Art und Weise der Gestaltung der Rückinformation oder von dem den Experten zur Verfügung gestellten Zeitfonds an dieser Stelle (noch) nicht betrachtet werden.

Folgenden operationalen Hypothesen zur Güte der Delphi-Urteile kann bei empirischen Tests innerhalb der **Expertenhypothesen** nachgegangen werden:

1. Die Güte der Urteile ist erfahrungsabhängig. Dies betrifft wiederum verschiedene Erfahrungsdimensionen:
  - die Häufigkeit des Umgangs mit statistischen Verteilungen,
  - die mit der eigenen Profession zusammenhängenden Erfahrungen,
  - die - damit jeweils spezifisch vermittelte - Vertrautheit mit dem zu beurteilenden, inhaltlichen Untersuchungsgegenstand sowie
  - die eigene Zugehörigkeit zur Untersuchungspopulation, d.h. hier konkret die Vertrautheit mit der untersuchten Altersgruppe der 13- bis 29-jährigen in Ost- und Westdeutschland.

Es wird angenommen, daß je ausgeprägter die jeweiligen Erfahrungen sind, desto größer der Erfolg der Schätzung ist.

---

1) Einen Ansatzpunkt für die Schätzung von Ergebnissen einer Bevölkerungsbefragung bieten auch methodische Forschungen zu Interviewfälschungen. "Trotz der potentiellen Dramatik und der gravierenden Konsequenzen ist das Problem der Interviewfälschung jedoch bislang kein nennenswertes Thema wissenschaftlicher Diskussion gewesen." (Reuband, 1990, S. 706) Immerhin liegen jedoch dazu einige empirische Tests vor, deren Ergebnisse auch in unserem Kontext einbezogen werden sollen.

2. Bei den Urteilen spielen kognitiv-psychologische Aspekte eine Rolle. So sind Unterschiede zu erwarten zwischen einerseits mehr rational-akribisch und andererseits eher emotional-spontan urteilenden Experten. Reuband (1990, S. 712) unterscheidet im Rahmen seiner Betrachtungen über Interviewfälschungen sogar vier Strategien, mit denen die Ergebnisse von Bevölkerungsbefragungen ermittelt werden können: (1.) die am häufigsten gewählte Orientierung an Stereotypen, (2.) die Vorstellung einer konkreten Person, (3.) zufälliges Antworten sowie (4.) die Nutzung mehrerer Strategien gleichzeitig.

Es wird vermutet, daß rational urteilende Experten mehr eigenes Hintergrundwissen aktivieren können und sich deshalb den Verteilungen besser annähern bzw. diese eher "treffen".

3. Einen wesentlichen Aspekt, der den Erfolg von Delphi-Studien bestimmt, stellt die Gruppendynamik dar. Je weiter die ursprüngliche Schätzung vom Gruppenmittel entfernt liegt und je höher das Vertrauen in die Fähigkeit der anderen Experten ist, desto größer ist die Bereitschaft, sich in den folgenden Wellen dem Gruppenmittel anzupassen. Je mehr von den einzelnen Experten bei ihren Urteilen auf die Meinung der Gruppe Wert gelegt wird, desto besser wird das Ergebnis der einzelnen Schätzung, vorausgesetzt, das Gruppenergebnis liegt dem tatsächlichen näher. Weiter wird angenommen, daß die verschiedenen Erfahrungsdimensionen jeweils ein unterschiedliches Gewicht für das Gelingen der Schätzungen haben. Dies trifft besonders für jene Experten zu, die ursprünglich nur vage strukturierte Vorstellungen von den zu schätzenden Verteilungen haben.

4. Je höher die eigene Kompetenz bei der abgegebenen Schätzung angesehen wird, desto geringer ist die Bereitschaft, sich im Urteil zu revidieren und desto eher wird das Gruppenergebnis ignoriert.

5. Den Experten ist es möglich, ihre eigene Kompetenz realistisch einzuschätzen. Je höher die Experten subjektiv ihre eigene Kompetenz einschätzen, desto genauer ist damit auch die von ihnen abgegebene Schätzung. Die selbsteingeschätzte Kompetenz steigt von Welle zu Welle.

Folgenden Vermutungen bezüglich der Güte der Delphi-Urteile kann prinzipiell innerhalb der **Inhaltshypothese** nachgegangen werden:

1. Die Abweichungen der Expertenschätzungen von den tatsächlichen Urteilen wird bestimmt durch die Verteilungsform der zu schätzenden Antworten.

Normalverteilte Antworten werden mit größerem Erfolg geschätzt als schief verteilte Antwortmuster, bzw. Mittelwerte, die mit einer geringen Streuung in Verbindung stehen, werden erfolgreicher geschätzt als solche, die mit einer breiten Streuung verbunden sind (vgl. auch Abschnitt 2.1.).

2. Die Güte der Schätzungen wird beeinflusst durch - von den Experten nicht erwartete - Extremwerte in den Originaldaten.

Bessere Schätzungen werden bei in der tatsächlichen Befragung weniger extrem beurteilten Inhalten abgegeben. Das Gleiche gilt für von den Experten nicht vorhersehbare Unterschiede zwischen ost- und westdeutschen Jugendlichen.

3. Es besteht eine Beziehung zwischen der Differenziertheit der zu schätzenden Ergebnisse (vier-, zweistufig bzw. "einstufig" Mittelwerte) und der Güte der Schätzung.

Am besten nähern sich die Schätzungen den tatsächlichen Werten bei den eher groben Mittelwertschätzungen an, am wenigsten gut bei den komplizierter und detaillierter zu schätzenden vierstufigen Antwortverteilungen. Dies gilt auch für die Treffsicherheit nach allen Delphi-Wellen.

Die genannten Hypothesen sollen im Rahmen dieses Delphi-Projektabschnittes einer ersten Bearbeitung unterzogen werden. Um solche Vermutungen präzisieren zu können, wurden - nach Abschluß aller Befragungswellen - ergänzende Informationen über die Experten teilstandardisiert erhoben. Diese betreffen analog zu den genannten Hypothesen:

- den bisherigen Umgang mit der untersuchten Thematik Zeitbewußtsein und Wichtigkeitsfragen;
- das dem von Welle zu Welle rückgemeldeten Gruppenergebnis beigemessene Gewicht;
- die den ZUMA-Experten insgesamt zugeschriebene Kompetenz, Verteilungen zu schätzen;
- die Erfahrungen mit dem Umgang mit statistischen Verteilungen;
- der Grad an Spontaneität beim Antworten in der ersten und in den folgenden Wellen;
- der Umgang mit bzw. Kontakte zu Jugendlichen;
- die für das Ausfüllen benötigte Zeitdauer pro Welle;
- den Beruf bzw. die Studienrichtung des Experten;
- die Berufserfahrung des Experten;
- alle weiteren während der Befragungswellen gewonnenen Eindrücke, die offen erfragt wurden.

## **5. Feldbericht**

Am 12. Januar 1994 wurde mit der ersten Welle der Delphi-Expertenbefragung am ZUMA begonnen. Dazu wurden ein Testbogen und ein dazu gehörendes Anschreiben ausgearbeitet (vgl. Anhang 1). Diese wurden zunächst an zwanzig ZUMA-Mitarbeiter mit dem geschilderten Ziel, die Ergebnisse einer Bevölkerungsbefragung zu schätzen, übergeben. Da es sich bei der Grundgesamtheit dieser Untersuchung um Jugendliche handelte, erschien es angebracht, auch fünf am ZUMA tätige studentische Hilfskräfte als Experten in den Test einzubeziehen. Da zwei der zunächst angesprochenen Mitarbeiter mit den Daten dieser Untersuchung vertraut waren bzw. als Projektleiter diese Untersuchung betreut hatten, konnten sie an dem Test nicht teilnehmen.

Deshalb wurden bereits nach kurzer Zeit zwei andere Teilnehmer nachrekrutiert. In der ursprünglich vorgesehenen Feldzeit von acht Tagen konnten damit zunächst nicht alle 20 Befragungen realisiert werden. Bis zum 27. Januar war dann der Rücklauf vollständig. Rückfragen der Teilnehmer, etwa das Anliegen der Befragung bzw. die Ausfüllregeln betreffend, traten kaum auf. Kommentare gab es in der Richtung, daß es sich bei einer solchen Art der Befragung um ein kompliziertes Anliegen handele.

Alle 20 ausgegebenen Fragebogen der ersten Welle konnten für die Auswertung verwandt werden. Antwortausfälle traten hier nur selten auf. 17 der zurückerhaltenen Fragebogen waren vollständig ausgefüllt. Noch am gleichen Tag wurden die Auswertung vorgenommen, die Ergebnisse rückgemeldet und damit die zweite Welle gestartet. Für die Auswertung wurde ein SPSS/PC-System-File erstellt, in dem die Namen der Teilnehmer anonymisiert wurden. Für die zweite Welle der Testbefragung wurde erneut ein eigenes Anschreiben ausgearbeitet (vgl. Anhang 2), in dem vor allem die Art der statistischen Rückinformation erklärt wurde. Ergänzt wurde am Schluß des Bogens eine Frage nach der empfundenen Sicherheit bei der Beantwortung der Fragen (vgl. Testbogen zur zweiten Welle im Anhang 2).

Der Rücklauf der zweiten Welle betrug wiederum 100% und war am 21. Februar abgeschlossen. Erneut waren fast alle Fragebogen vollständig beantwortet. Achtzehn der zurückerhaltenen Fragebogen waren komplett ausgefüllt. Auch diesmal wurde noch am selben Tag die Auswertung vorgenommen und mit der dritten Welle begonnen. Für die Auswertung wurde das bestehende SPSS/PC-System-File um die Daten der zweiten Welle ergänzt. Für die dritte Welle wurde wiederum ein kurzes eigenes Anschreiben erarbeitet (vgl. Anhang 3). Die statistische Rückinformation über die Ergebnisse des vorangegangenen Tests erfolgte analog zur vorigen Welle.

Nachdem uns die einzelnen ausgefüllten Fragebogen der dritten Welle zugestellt worden waren, begann jeweils die abschließende ergänzende Befragung der Experten. Diese vierte Erhebung (vgl. Anhang 4) begann bereits am 23. Februar.

Aus mündlichen Äußerungen einiger (konkret von drei) Experten ging hervor, daß das gewählte Vorgehen, bei der Rückmeldung neben dem Gruppenergebnis jeweils auch die eigenen Schätzungen der vorangegangenen Welle mitzuteilen, nicht als günstig empfunden wurde. Es wurde vorgeschlagen, die Experten jeweils "völlig neu", d.h. ohne Möglichkeit einer Bezugnahme auf die eigene vorangegangene Antwort, schätzen zu lassen. Andererseits wurde auch die gegenteilige Auffassung vertreten, nach der die zurückgegebenen Erhebungsbögen aus der vorangegangenen Welle besonders nützlich für die erneute Meinungsbildung waren. Weiterhin gab es Hinweise, daß mit drei Wellen und abschließender Zusatzbefragung die Teilnahmebereitschaft der Experten wahrscheinlich ausgeschöpft war.

Die Befragung der dritten Welle war am 23. März abgeschlossen, die Erhebung der zusätzlichen Informationen zu den Eindrücken der Experten dauerte bis zum 30. März 1994.



## 6. Auswertung

### 6.1. Allgemeines Konzept

Die folgende Darstellung der Ergebnisse der Expertenschätzungen beinhaltet deren Beurteilung aufgrund verschiedener Kriterien <sup>2)</sup>:

1. Die Bestimmung der Güte der bei jeder einzelnen Welle für *jede einzelne Frage* abgegebenen Schätzungen für die Gesamtheit des Expertengremiums.
2. Die Bestimmung der Güte der Einzelurteile *jedes Experten* in jeder Welle.
3. Sowohl die Gruppierung der Experten (nach dem Erfolg ihrer Schätzungen) als auch eine Gruppierung der geschätzten Items (nach der Güte der Schätzungen).
4. Schließlich die Kopplung beider Aspekte derart, daß Aussagen darüber möglich werden, welchen Experten es gelingt, welche Indikatoren erfolgreich zu schätzen.
5. Die Konfrontation der dabei gebildeten Gruppen mit den - expertenseitig vor allem in der Zusatzbefragung gewonnenen und indikatoreseitig aus den Originalverteilungen entnommenen - Erklärungsgrößen.

In der Darstellung werden zunächst (6.2.) die Resultate der vierstufigen Verteilungsschätzungen präsentiert, im folgenden Abschnitt (6.3.) werden die Ergebnisse der Schätzungen der Alternativantworten referiert, schließlich (6.4.) sollen die bei den Indikatoren zu den Wichtigkeitsurteilen (Mittelwertfragen) abgegebenen Schätzungen beurteilt werden.

Wir halten diese Abfolge der Präsentation der Ergebnisse entsprechend der verschiedenen Zielstellungen bei den Schätzungen für sinnvoll, weil sie jeweils sehr unterschiedliche inhaltliche und methodische Anforderungen an die Experten beinhalten und infolge dessen auch zu qualitativ recht unterschiedlichen Ergebnissen geführt haben.

Abschließend (6.5) stellen wir die Ergebnisse der zusätzlichen Befragung (vierte Welle) zu den Eindrücken der Experten bezüglich unseres Tests dar.

### 6.2. Auswertung der Indikatorenschätzungen bei vierstufigen Skalen<sup>3)</sup>

In einem ersten Auswertungsschritt werden die in den drei Wellen erzielten Ergebnisse der Schätzungen vierstufiger Antwortverteilungen gegenübergestellt. Dabei werden zunächst jeweils der Mittelwert (arithmetisches Mittel) und die Streuung (Interdezilbereich) für jede einzelne erfragte Antwortstufe betrachtet. Auf diese Weise wird ein erster Überblick über die gewonnenen Ergebnisse ermöglicht.

---

2) Dieses Konzept konnte im vorliegenden ersten Bericht nicht vollständig umgesetzt werden. Weitere Arbeiten zur Auswertung der Delphi-Testerhebung haben zu folgen.

3) Dieser Abschnitt entstand unter Mitarbeit von Birgit Rimmelspacher.

Dabei wird deutlich (vgl. Tabelle 1), daß sich die bei den Schätzungen einer vierstufigen Antwortverteilung (Fragen 1 bis 5) erzielten einzelnen Mittelwerte auf die wahren Werte der Shell-Studie zubewegen, zugleich verkleinert sich von Welle zu Welle die Streuung der Urteile.

**Tabelle 1:**

*Veränderungen der Mittelwerte (arithmetische Mittel - MW - in Prozent) und der Streuungen (Interdezilbereiche - IDB - in Prozent) der Schätzungen aller Experten von Welle zu Welle bei den fünf Fragen mit vierstufigem Antwortmodell*

	1.Welle		2.Welle		3.Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
<b>1.Frage</b>							
überh. nicht	20	10-30	18	15-25	18	10-25	14
weniger gut	31	20-40	33	20-40	32	25-40	36
gut	30	13-50	32	20-40	33	25-40	35
sehr gut	19	5-25	17	10-20	17	10-20	15
<b>2.Frage</b>							
überh. nicht	15	5-20	14	10-20	14	9-20	9
weniger gut	34	20-50	33	25-40	32	25-40	40
gut	33	20-50	35	30-50	36	27-40	37
sehr gut	18	10-20	18	10-25	18	10-20	14
<b>3. Frage</b>							
überh. nicht	10	0-18	9	5-20	8	5-13	3
weniger gut	21	10-30	20	10-30	20	10-28	20
gut	40	20-60	44	30-50	44	35-50	49
sehr gut	29	15-50	27	15-40	28	20-35	28
<b>4.Frage</b>							
überh. nicht	29	15-40	27	20-40	28	20-40	17
weniger gut	41	30-50	42	30-50	41	30-50	34
gut	20	10-30	20	12-25	20	15-25	37
sehr gut	10	5-20	11	5-20	11	5-20	12
<b>5.Frage</b>							
überh. nicht	26	10-40	21	10-30	20	10-30	32
weniger gut	35	20-45	37	30-45	38	30-45	47
gut	24	15-40	27	20-40	28	20-40	17
sehr gut	15	3-20	15	10-20	14	10-20	4

In den meisten Fällen (13 von 20) liegt der wahre Wert innerhalb des Interdezilbereiches der letzten Schätzung. Die in der dritten Welle geschätzten Verteilungsformen kommen bei einer



optischen Gegenüberstellung den wirklichen Verteilungsformen mehrheitlich nahe (vgl. dazu die Abbildungen 1 bis 5).

Um eine präzisere Aussage über die Abweichungen der Verteilungen der Expertenschätzungen von denen aus der Shell Jugendstudie treffen zu können, wurde ein  $\chi^2$ -Test auf Gleichheit der Verteilungsformen berechnet. Dieser Test soll im folgenden ausführlich dokumentiert werden, weil die ermittelten Prüfgrößen für mehrere Analysen genutzt werden.

Für den  $\chi^2$ -Test wurde als Irrtumswahrscheinlichkeit 5 % gewählt, damit ergibt sich bei drei Freiheitsgraden für die Prüfgrößen ein kritischer Wert von 7.81.<sup>4)</sup>

Nehmen nun die mit Hilfe des  $\chi^2$ -Tests ermittelten Prüfgrößen für die Verteilungen der Expertenschätzungen diesen kritischen Wert an oder überschreiten ihn, so ist davon auszugehen, daß die Verteilung der Expertenschätzung und die aus der Shell Jugendstudie nicht identisch sind, d.h. die Nullhypothese zurückzuweisen ist.

Obwohl dieser kritische Wert für alle Fragen überschritten wurde, und die Nullhypothese von daher abgelehnt werden muß, zeigt sich, daß sich die Schätzungen der Experten in vier von fünf Fragen den wahren Verteilungen angenähert haben und der Test insofern durchaus erfolgreich verlaufen ist. Tabelle 2 gibt eine Übersicht über die ermittelten Prüfgrößen für alle Experten.

---

4) Die Berechnung der Prüfgrößen für jede einzelne Frage in jeder einzelnen Welle für jeden Experten ruft in einigen wenigen Fällen praktische Schwierigkeiten hervor, da es Experten gibt, die vermuteten, daß bei bestimmten Fragen von den befragten Jugendlichen manche Antwortmöglichkeiten nicht ausgewählt wurden, so daß in diesen Fällen der Wert Null vergeben wurde. Dadurch ist eine Berechnung der Prüfgröße nicht mehr ohne weiteres möglich. Es reicht allerdings auch nicht aus, daß alle Schätzwerte verschieden von Null sind. Zusätzlich ist noch erforderlich, daß kein Schätzwert geringer als 3/40 (gerundet in Prozent) ist, da eine unbedenkliche Anwendung des  $\chi^2$ -Anpassungstests nur dann möglich ist, wenn alle absoluten erwarteten Häufigkeiten größer als 1 sind ( $np_i > 1$  für alle  $i$ ) und weniger als ein Fünftel kleiner als 3 sind (vgl. Lohse/Ludwig 1982, S.115). In der Regel hilft man sich in Fällen, in denen die Werte dieser Anforderung nicht genügen, indem "benachbarte", d.h. ähnlich lautende Antwortmöglichkeiten zusammengefaßt werden (vgl. Storm 1979, S.187). In der am ZUMA durchgeführten Untersuchung kommt es jedoch auch vor, daß ein Experte die Schätzung abgab, daß lediglich eine der vier möglichen Antworten von den Jugendlichen ausgewählt wurde. Somit müßten bei der üblichen Vorgehensweise alle Schätzwerte dieses Experten eine bestimmte Frage betreffend zu einem Wert zusammengefaßt werden, was zur Folge hätte, daß sich für die Prüfgröße der Wert Null errechnen würde, woraus falsche Schlüsse gezogen werden könnten. Zur Lösung dieses Problems wurden generell die geschätzten Nullwerte durch den Wert 3/40 ersetzt und alle übrigen Schätzwerte proportional verringert.

Abbildungen 1 bis 5:

Von den Experten geschätzte sowie Original-Verteilungen aus der Shell Jugendstudie

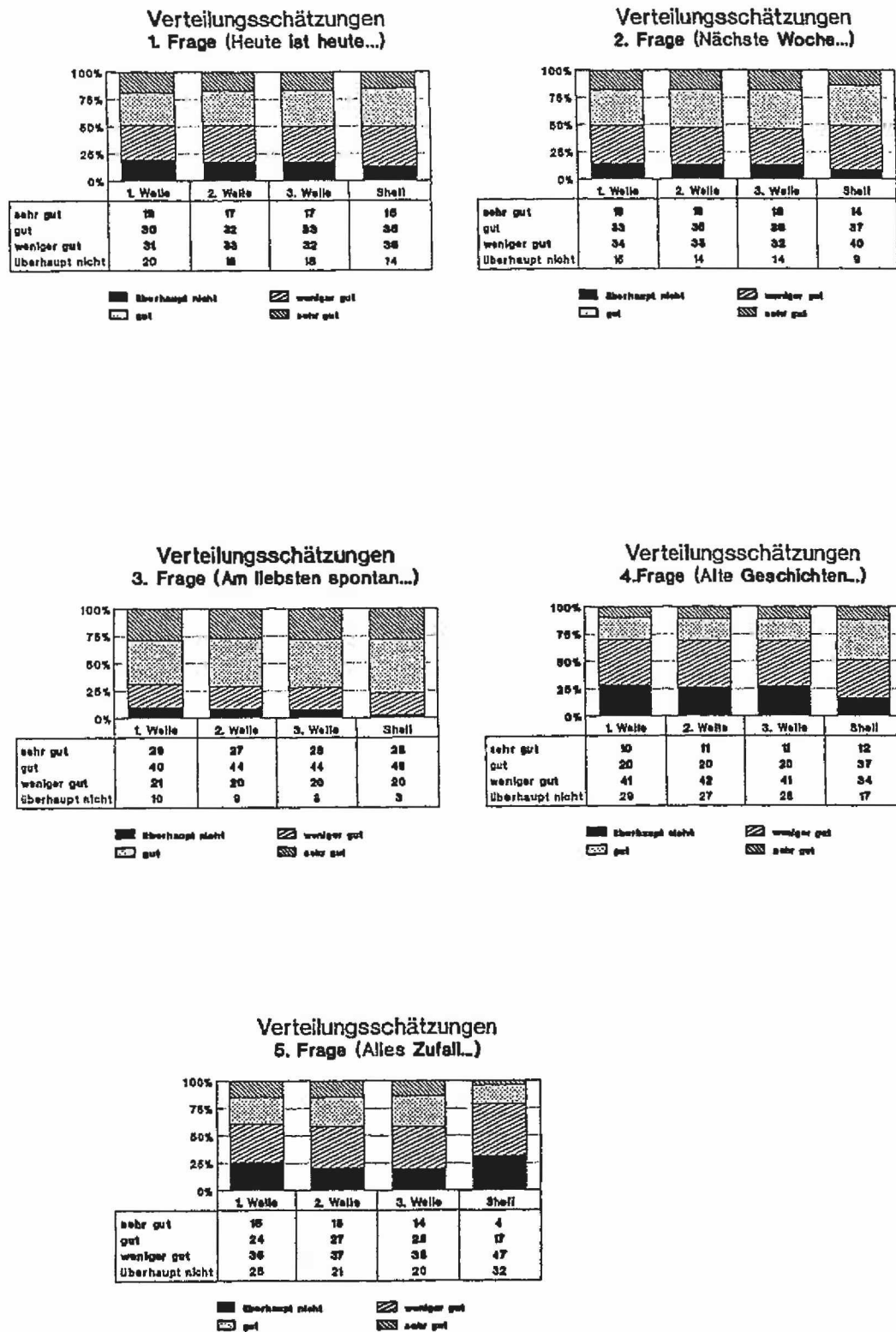


Tabelle 2:  
*Prüfgrößen des  $\chi^2$ -Tests der Expertenschätzungen für die fünf Fragen mit vierstufigem Antwortmodell für alle Experten ("Gesamtexperte"<sup>5</sup>)*

Fragenr.	1. Welle	Prüfgröße 2. Welle	3. Welle
1	176.34	112.72	76.96
2	196.56	186.94	195.67
3	279.24	200.66	169.04
4	847.27	860.26	784.49
5	610.95	850.33	893.55

Bei den Fragen 1, 3 und 4 hat sich von Welle zu Welle eine kontinuierliche Verbesserung des Schätzergebnisses eingestellt. Insbesondere die Frage 1 kann als außerordentlich erfolgreiche Schätzung des in der Shell Jugendstudie ermittelten Ergebnisses angesehen werden. Auch bei Frage 2 liegt die in der dritten Welle geschätzte Verteilung näher an der tatsächlichen als die der ersten Welle. Bei Frage 5 hat sich hingegen die Prüfgröße von Welle zu Welle erhöht, die finale Schätzung liegt im Vergleich der fünf Fragen am weitesten von der tatsächlichen Verteilung entfernt. Dies läßt sich daraus erklären, daß in der 2. und 3. Welle die Antworthäufigkeiten fast vollständig von allen Experten einheitlich entweder über- oder unterschätzt wurden, während in der 1. Welle beispielsweise die Häufigkeit der 1. Antwortmöglichkeit zwar auch von den meisten Experten unterschätzt, jedoch auch von zwei Experten ganz erheblich überschätzt wurde (Ausreißer), so daß die Gesamtexpertenschätzung dem wahren Wert deutlich näher kommt und deshalb die für die erste 1. Welle ermittelte Prüfgröße deutlich geringer ist.

Die Verteilung von Frage 5 ist von den fünf aus der Shell Jugendstudie ausgewählten zugleich die mit der größten Schiefe, sie ist stark linkssteil ( $Sch=0.617$ ). Die anderen Verteilungen der Originaldaten sind dagegen entweder (nahezu) symmetrisch (1. Frage:  $Sch=0.005$ ; 2. Frage  $Sch=0.057$ ; 4. Frage:  $Sch=-0.012$ ) bzw. für die dritte Frage ( $Sch=-0.413$ ) ausgeprägt rechtssteil. Aus diesem Ergebnis allerdings abzuleiten, daß bestimmte Verteilungsformen besser oder schlechter geschätzt werden können, ist nicht angemessen. Hierzu bedarf es weiterer Tests.

Im folgenden sollen die Schätzleistungen der einzelnen Experten beurteilt werden <sup>6)</sup>. Dabei kann zunächst für jede der von den 19 Experten <sup>7)</sup> abgegebenen Schätzungen registriert werden, ob die finale Schätzung besser (Prüfgröße verringert) oder schlechter (Prüfgröße erhöht) als die der ersten Welle war oder ob die Schätzung nicht verändert wurde (Prüfgröße konstant). Bei 54 der

5) Unter "Gesamtexperte" verstehen wir eine synthetische Größe, die gebildet wurde aus den mittleren Expertenschätzungen der einzelnen Antwortstufen über alle Fragen. Für diese wurde dann die  $\chi^2$ -Prüfgröße berechnet.

6) Diese können hier nicht im einzelnen dargestellt werden.

7) Ein Experte verweigerte die Schätzung dieser Antworten in allen drei Wellen.

95 so zu beurteilenden Schätzungen (5 Fragen x 19 Experten) verbesserte sich das Ergebnis von der ersten zur dritten Welle, bei 32 Schätzungen blieb das Urteil konstant und bei nur neun Schätzungen lag die in der dritten Welle ermittelte Prüfgröße über der der ersten Welle.

Pro Experte ergaben sich die folgenden mittleren Schätzfehler (durchschnittliche Prüfgrößen pro Experte) für die fünf Fragen nach den Antwortverteilungen in der Shell Jugendstudie (vgl. Tabelle 3).

**Tabelle 3:**  
*Schätzfehler (durchschnittliche  $\chi^2$ -Prüfgrößen für die fünf Fragen mit vierstufigem Antwortmuster) pro Experte und Welle*

Experte- Nr.:	Durchschnittliche Prüfgröße		
	1. Welle	2. Welle	3. Welle
1	2448.857	620.949	540.315
2	2879.929	1376.131	1376.131
3	-	-	-
4	1772.723	1772.723	1772.723
5	59496.108	935.012	535.602
6	1428.383	1784.828	1775.178
7	328.610	289.828	414.517
8	558382.519	454.756	798.828
9	2590.391	1355.855	870.080
10	681.406	564.322	490.265
11	1421.895	471.455	406.809
12	2219.331	2219.331	1513.330
13	1145.748	547.483	451.688
14	2311.486	2311.486	2311.486
15	3819.532	608.028	388.737
16	15762.093	15762.093	15762.093
17	1202.568	674.736	581.686
18	947.463	495.645	506.554
19	654.777	507.878	509.671
20	770.922	755.665	394.726

Es ist bei 15 der 20 Experten festzustellen, daß sich die (mittleren) Verteilungen in der 3. Welle nicht mehr in dem Ausmaße von den gegebenen Verteilungen unterscheiden wie in der 1. Welle, die Schätzungen sich also verbessert haben. Nur bei zwei Experten entfernten sich die geschätzten (mittleren) Verteilungen von den Originalverteilungen der Shell Jugendstudie.

Insgesamt ergeben sich für diesen Aufgabenteil, das Schätzen von vierstufigen Antwortverteilungen, drei Tendenzen:

1. Mehrheitlich verbessern sich die Schätzleistungen der Experten von Welle zu Welle.

2. Die Schätzungen der Experten insgesamt sind tendenziell von höherer Qualität als die einzelner Experten.
3. Es gelingt allerdings nicht, die den strengen Kriterien eines  $\chi^2$ -Tests entsprechende Identität der geschätzten und der empirischen Verteilungsformen zu erzeugen.

### 6.3. Auswertung der Indikatorenschätzungen bei zweistufigen Antwortskalen

Auch die Ergebnisse der insgesamt vier Schätzungen von alternativen Antwortverteilungen (Fragen 6 bis 9, vgl. Tabelle 4) können insgesamt als erfolgreich bezeichnet werden.

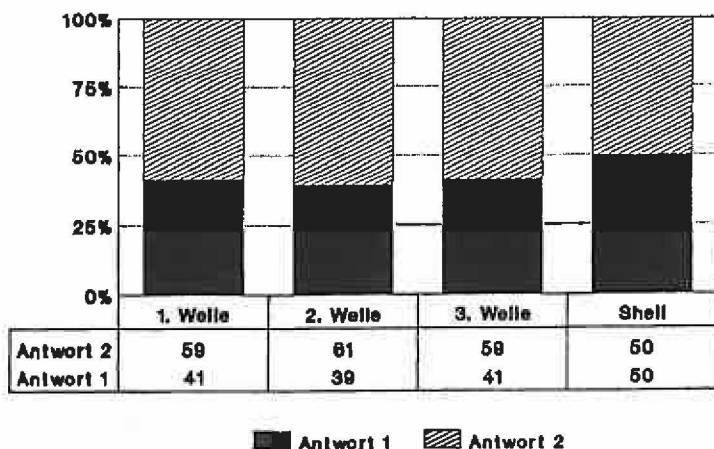
**Tabelle 4:**

*Veränderungen der Mittelwerte (arithmetische Mittel - MW -in Prozent ) und der Streuungen (Interdezilbereiche - IDB - in Prozent) der Schätzungen der Experten von Welle zu Welle bei den vier Fragen mit zweistufigen Antwortvorgaben im Vergleich zu den Ergebnissen der Shell-Studie*

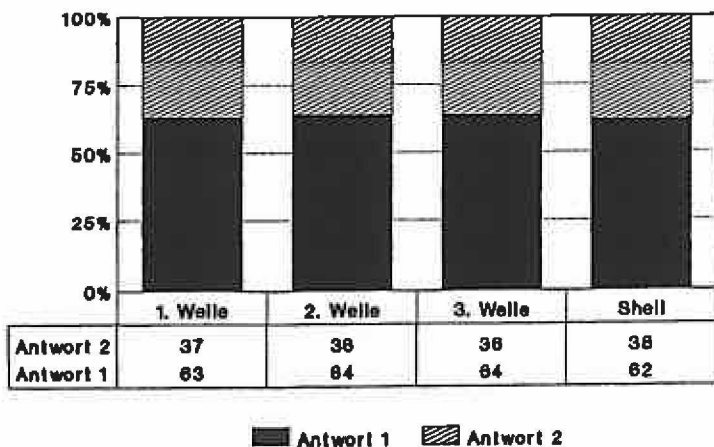
	1.Welle		2.Welle		3.Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
<b>6. Frage</b>							
Antwort 1	41	30-60	39	30-50	41	30-50	50
Antwort 2	59	40-70	61	50-70	59	50-70	50
<b>7. Frage</b>							
Antwort 1	49	30-70	47	40-60	48	40-55	37
Antwort 2	51	30-70	53	40-60	52	45-60	63
<b>8. Frage</b>							
Antwort 1	63	30-80	64	50-75	64	50-75	62
Antwort 2	37	20-70	36	25-50	36	25-50	38
<b>9. Frage</b>							
Antwort 1	50	30-70	50	40-65	48	40-60	43
Antwort 2	50	30-70	50	35-60	52	40-60	57

Als Kriterien für die Beurteilung der Schätzungen werden zunächst wieder die Interdezilbereiche sowie die Mittelwerte betrachtet. So verringern sich die Interdezilbereiche (erste Welle im Durchschnitt 40 Prozentpunkte, zweite Welle 23 Prozentpunkte und dritte Welle 20

## Alternativschätzungen 6. Frage

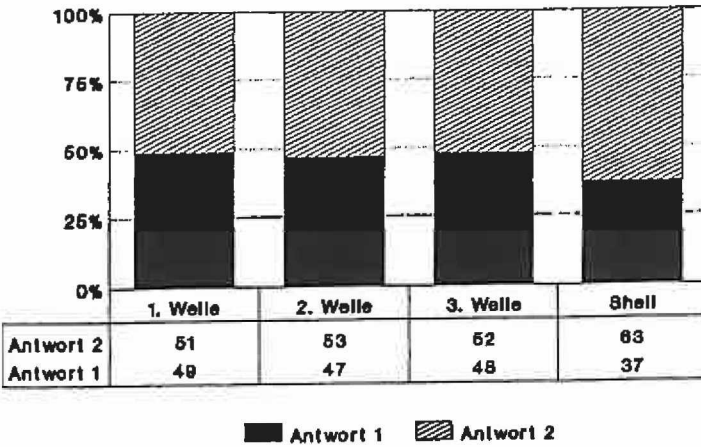


## Alternativschätzungen 8. Frage

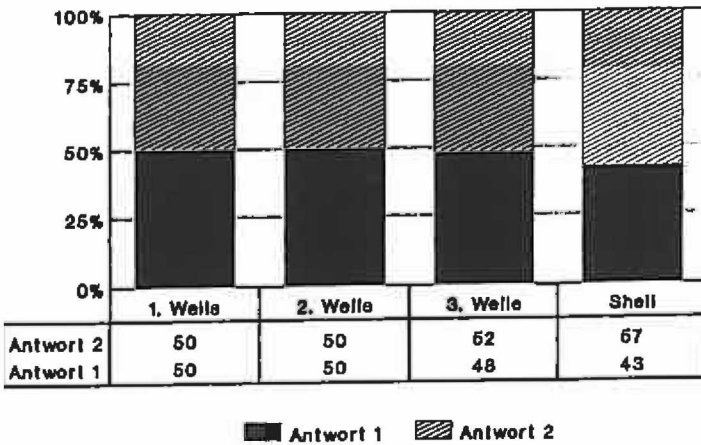


Abbildungen 6 bis 9:  
 Von den Experten geschätzte sowie Original-Verteilungen aus der Shell Jugendstudie für die  
 Alternativfragen

### Alternativschätzungen 7.Frage



### Alternativschätzungen 9.Frage



Prozentpunkte) und bewegen sich auf die wahren Antworten zu. Diese liegen - mit Ausnahme der Frage 7 - nach der dritten Welle jeweils innerhalb dieses Bereichs. Die 7. Frage betraf die Gestaltung des Lebens entweder orientiert an der eigenen Individualität (= Antwort 1) oder ausgerichtet an den Erwartungen der anderen Menschen (= Antwort 2). Dabei haben die Experten die erste Vorgabe überschätzt und dementsprechend die zweite unterschätzt. D.h., es wurde von ihnen angenommen, daß die Jugendlichen 1992 geringeren Wert auf die Gemeinsamkeiten mit anderen legten, als dies wirklich der Fall war.

Um nach weiteren Aussagen über mögliche Ursachen für die unterschiedlichen Schätzerfolge zu suchen, werden in der folgenden Tabelle 5 die Differenzen zwischen den Ergebnissen der Schätzungen in der letzten Welle und den tatsächlichen Ergebnissen als Ausdruck für die jeweils bei den Schätzungen begangenen Fehler (der Gesamtpopulation sowie getrennt nach Befragten in Ost- und Westdeutschland) gezeigt.

**Tabelle 5:**  
*Schätzfehler (Abstände zwischen den Mittelwerten der Expertenschätzungen ( $N_E=20$ ) in der dritten Welle und den wahren Werten in der Shell Jugendstudie, unterschieden nach Gesamtheit der befragten Jugendlichen ( $N_{SHELL}=4005$ ) sowie Befragten in Ost- ( $n=1336$ ) und Westdeutschland ( $n=2669$ ), Angaben in Prozentpunkten; wobei (I) auf eine Schätzung innerhalb und (A) auf eine Schätzung außerhalb des Interdezilbereichs verweisen)*

Gesamtpopulation		Westen	Osten
<b>6. Frage</b>	9 (I)	5 (I)	17 (A)
<b>7. Frage</b>	11 (A)	8 (I)	17 (A)
<b>8. Frage</b>	2 (I)	2 (I)	4 (I)
<b>9. Frage</b>	5 (I)	3 (I)	9 (A)

Hier wird deutlich, daß die (zufällig) ausnahmslos westdeutschen ZUMA-Experten die Unterschiede zwischen den Befragten in West- und Ostdeutschland nicht adäquat einzuschätzen vermochten. So wurde folgerichtig die Frage 8, bei der kein Unterschied zwischen Ost und West in den zu schätzenden Urdaten bestand, am besten geschätzt. Nur in diesem Fall würde eine Übertragung der Schätzung auch für die Subpopulation der ostdeutschen Jugendlichen erfolgreich sein. Weiterhin ist auffällig, daß die Schätzungen für alle vier Indikatoren innerhalb der jeweiligen Interdezilbereiche liegen, wenn man lediglich die Befragungsergebnisse der westdeutschen Jugendlichen als zu schätzende zugrunde legt. Demgegenüber verschlechtert sich das Ergebnis beträchtlich, wenn nur die ostdeutschen Befragten betrachtet werden: in drei von vier Fällen liegt hier der finale Schätzwert außerhalb des Interdezilbereiches.



Zwei Schlußfolgerungen können daraus zunächst gezogen werden:

Dieses Ergebnis könnte darauf hindeuten, daß Schätzungen von Ergebnissen von Bevölkerungsbefragungen erschwert werden, wenn in den Originaldaten starke Unterschiede zwischen (jeweils relativ großen) Subpopulationen auftreten.

Zum anderen werden offenbar von den Experten tatsächlich für die Schätzungen die eigenen Erfahrungen herangezogen und nicht - wie teilweise bei "non-attitudes" - eine imaginäre Zufallsziehung<sup>8)</sup> angestellt, um Unwissen zu verbergen und trotzdem zu den erfragten Sachverhalten eine Einschätzung abzugeben. Dies stellt zugleich einen wichtigen Hinweis für das prinzipielle Funktionieren der Delphi-Technik dar.

Weiterhin ist es für den Erfolg der Schätzungen zweistufiger Antwortskalen - entgegen unseren ursprünglichen Annahmen - unerheblich, wie stark jeweils die beiden Antwortvorgaben in den Urdaten besetzt sind.

Verschiedene weitere Hypothesen (vgl. Abschnitt 4.) lassen sich aufgrund der relativ geringen Anzahl einbezogener Experten mit diesem Test noch nicht weiter bearbeiten. Dies gilt beispielsweise für die Fähigkeit der Experten, die Qualität der eigenen Urteile adäquat einzuschätzen bzw. für die die Gruppendynamik beim Delphi betreffenden Vermutungen.

Auch für diese Schätzaufgabe sollen wieder die Erfolge der einzelnen Experten beurteilt werden. Tabelle 6 gibt dazu einen Überblick.

Demnach haben die Schätzungen von vier Experten (Nr. 2, 4, 14 und 16) in allen Wellen den gleichen Abstand zum wahren Wert, da diese Experten ihr Urteil im Verlauf des Tests nicht geändert haben. Bei einer Reihe anderer Experten hat sich die Schätzung in der dritten Welle deutlich gegenüber der ersten Schätzung verbessert, in einem Fall (Nr. 20) sogar um fast 50 Prozentpunkte. Verschlechterungen von Schätzungen traten zwischen der ersten und der zweiten Welle in keinem Fall auf. Zwischen zweiter und dritter Welle gab es einige Verschlechterungen, insgesamt jedoch zeigt der Vergleich zwischen erster und dritter Welle ausschließlich bessere (bzw. zumindest gleichbleibende) Schätzerfolge. Gerade die Verringerung der mittleren Schätzfehler von Welle zu Welle (letzte Zeile der Tabelle 6) verdeutlicht noch einmal den Erfolg des Delphi-Ansatzes bei der Schätzung dieser vier Fragen.

---

8) Vgl. Abschnitt 2.1. dieses Berichts.

Tabelle 6:

*Schätzfehler (Summe der Beträge der Abstände der Schätzungen zu den wahren Werten in den einzelnen Wellen) pro Experte für die vier Alternativfragen, Angaben in Prozentpunkten, sowie Darstellung der Veränderung der Schätzfehler zwischen der ersten und der dritten Welle, positive Werte verweisen dabei auf eine Verringerung des Schätzfehlers.*

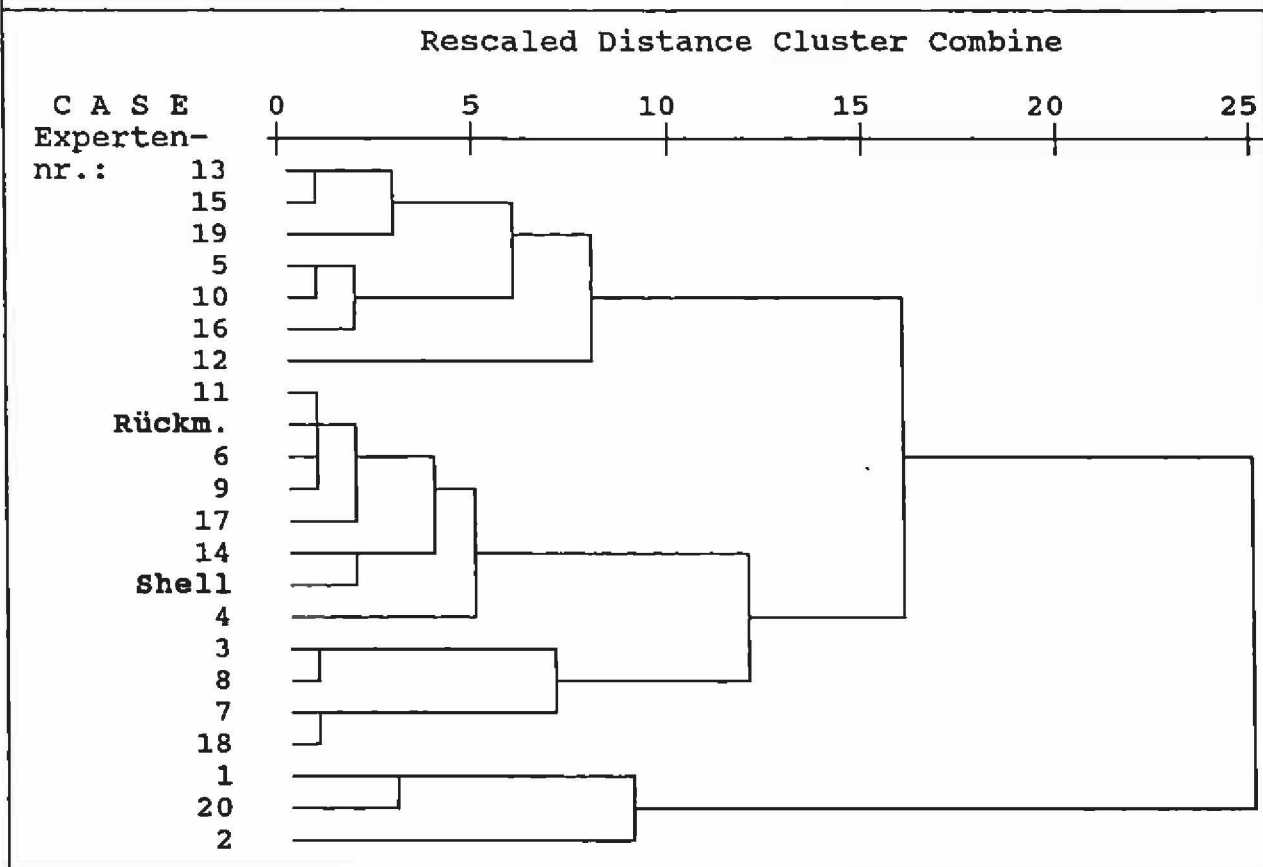
Experte Nr.	1. Welle	2. Welle	3. Welle	Diff. 1.-3. Welle
1	82	38	48	34
2	78	78	78	0
3	29	29	24	5
4	34	34	34	0
5	34	24	28	6
6	52	28	20	32
7	68	48	48	20
8	–	33	34	–
9	17	14	17	0
10	48	48	38	10
11	78	52	32	46
12	88	88	70	18
13	93	58	53	40
14	26	26	26	0
15	72	51	52	20
16	43	43	43	0
17	58	58	28	30
18	59	49	39	20
19	78	68	68	10
20	92	58	43	49
Mittelwert	59	46	41	18

Für die Darstellung der Struktur der Experten hinsichtlich ihrer unterschiedlichen Schätzerfolge für die vier Alternativfragen wurden Clusteranalysen berechnet. Als Merkmalsvariable fungierten die in der dritten Welle jeweils für die erste Antwortmöglichkeit der vier Indikatoren abgegebene Schätzung. Zunächst wurden als Anhaltspunkte für die Einschätzung der Lage der Cluster auch die Rückmeldung nach der zweiten Welle (Rückm.) sowie die Ergebnisse der Shell Jugendstudie (Shell) einbezogen. Das Dendrogramm, das als Ergebnis des Fusionierungsalgorithmus nach Ward entstanden ist, zeigt Abbildung 10.

Aufgrund des optischen Eindrucks scheint es zunächst plausibel, von der Existenz dreier Gruppen von Experten auszugehen. Dabei ergibt sich die folgende Zuordnung der Experten zu den Clustern:

Abbildung 10:

Dendrogramm nach der Ward Methode (Expertenschätzungen der dritten Welle sowie Shell-Werte und nach der zweiten Welle rückgemeldete Werte)



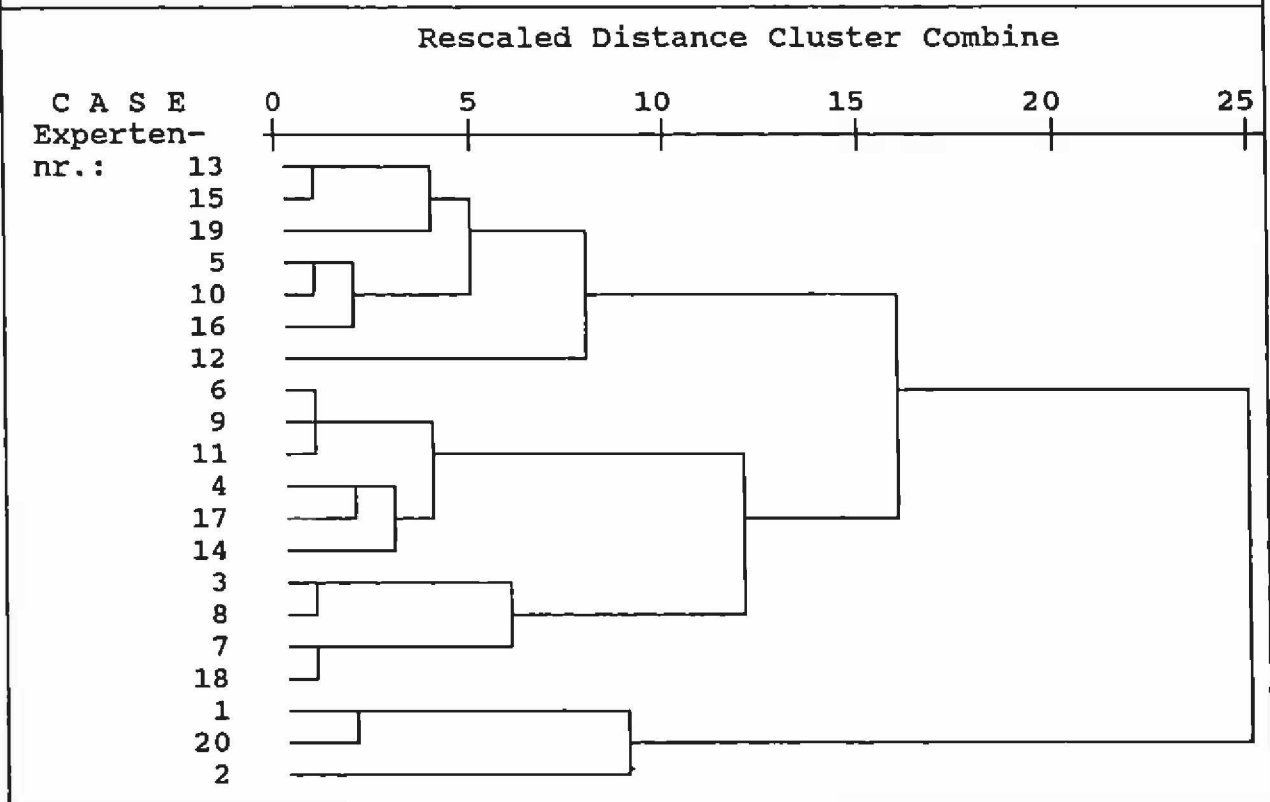
Die Experten mit den Nummern 11, 6, 9, 17, 14, 4, 3, 8, 7 und 18 befinden sich - gemeinsam mit den ("wahren") Shell-Werten sowie den Daten der Rückmeldung aus den Ergebnissen der Schätzungen der zweiten Welle - in einem Cluster I.

Ein weiteres Cluster II wird durch die Experten 13, 15, 19, 5, 10, 16 und 12 gebildet. Der Mittelpunkt dieses Clusters unterscheidet sich signifikant von dem des Clusters I, das aufgrund der Zuordnung der wahren Werte als das der Experten mit den besten Schätzerfolgen bezeichnet werden kann. Darüber hinaus existiert ein drittes Cluster III mit den Experten 1, 20 und 2, deren Schätzleistung offenbar stark von der der anderen Experten abweicht.

Als Erfolg der Schätzung insgesamt kann im übrigen angesehen werden, daß sich die aufgrund der Expertenschätzungen der vorangegangenen Welle rückgemeldeten und die wahren Werte in einem Cluster befinden.

Der Eindruck des Bestehens dreier Cluster bestätigt sich, wenn die Analyse ohne die Werte der Shell-Studie sowie die nach der zweiten Welle rückgemeldeten Werte wiederholt wird (vgl. Dendrogramm in Abbildung 11).

Abbildung 11:  
Dendrogramm nach der Ward Methode (Expertenschätzungen der dritten Welle)



Es könnte somit die Hypothese aufgestellt werden, daß es zunächst bezüglich des Schätzens von Alternativfragen Expertengruppen gibt, die diese Aufgabe mit unterschiedlicher Qualität bewältigten. Dabei wollen wir die Experten, deren Schätzungen sich in dem Cluster I mit den "wahren" Werten befinden, als die erfolgreichen Schätzer bezeichnen.

Bei den beiden anderen Expertengruppen handelt es sich offenbar um solche, die die wahren Werte jeweils in ähnlicher Weise über- oder unterschätzt haben.

Es soll nun mit Hilfe einer Diskriminanzanalyse festgestellt werden, ob sich die über die Clusteranalyse gewonnene Einteilung der Experten in mehr und weniger erfolgreiche unterstützen läßt.

Da die Richtung des Schätzfehlers (wahrer Wert über- oder unterschätzt) für die Beurteilung der Güte der Schätzung keine Rolle spielt, werden als diskriminierende Variablen die Beträge der Abweichungen der Schätzungen von den wahren Werten pro Experte eingesetzt.

Als erfolgreiche Experten (Expertengr. 1) werden - wie schon vereinbart - die des oben ermittelten Clusters I, als weniger erfolgreiche Experten (Expertengr. 2) die der Cluster II und III angesehen. Daraus ergibt sich direkt die empirische Verteilung der Zielvariablen.

Bei einem ersten Analyselauf konnte zwar eine mit  $p = .0057$  signifikante Diskriminanzfunktion ermittelt werden. Bei der aufgrund dieser Funktion ermittelten Klassenzugehörigkeit traten jedoch

zwei von den Ergebnissen der Clusteranalyse abweichende Eingruppierungen auf (für die Experten 5 und 20 wurde jeweils eine Aufnahme in die Gruppe der besseren Schätzer ermittelt). Nach der Korrektur der Expertenklassifizierung in diesen beiden Fällen wurde erneut eine Diskriminanzanalyse ausgeführt. Sie lieferte die folgenden Ergebnisse:  
Zunächst werden die mittleren Beträge der Abweichungen der Expertenschätzungen von den wahren Werten pro Indikator für die beiden Expertengruppen ausgewiesen (vgl. Tabelle 7).

Tabelle 7: Mittelwerte der Beträge der Abweichungen der Expertenschätzungen von den wahren Werten für die Alternativfragen				
	6. Frage	7. Frage	8. Frage	9. Frage
Expertengr.1	9.17	10.92	6.33	4.67
Expertengr.2	12.13	12.00	13.38	18.75

Es zeigt sich, daß tatsächlich bei allen Indikatoren die durchschnittlichen Schätzfehler in der ersten Expertengruppe geringer sind als in der zweiten. Besonders prononciert tritt der unterschiedliche Schätzerfolg in beiden Gruppen bei den Fragen 8 und 9 hervor.  
Die ermittelte Diskriminanzfunktion ist mit einem  $\chi^2$ -Wert von 29.23 bei 4 Freiheitsgraden hochsignifikant.  
Wie die folgende Tabelle 8 zeigt, besteht bei der aufgrund der Diskriminanzfunktion ermittelten und der ursprünglichen Klassifikation der Experten vollständige Übereinstimmung.

Tabelle 8: Klassifikations-Ergebnisse				
Tatsächliche Gruppe		Zahl der Experten	Geschätzte Gruppenzugehörigkeit	
			1	2
Gruppe	1	12	12 100%	0 0%
Gruppe	2	8	0 0%	8 100%
Anteil der korrekt klassifizierten Experten: 100%				

Damit besteht im weiteren die Möglichkeit zu prüfen, ob die ermittelte Klassifizierung der Experten in solche mit erfolgreichen und weniger erfolgreichen Schätzungen der Alternativfragen auch bei den inhaltlich und methodisch anders gearteten Mittelwertfragen auftritt.

#### 6.4. Ergebnisse der Mittelwertschätzungen bei Indikatoren mit siebenstufigen Antwortskalen

Im folgenden soll die Schätzung der Mittelwerte der 18 Indikatoren zu den Wichtigkeitsurteilen ausgewertet werden. Eine Übersicht über die von den Probanden im Durchschnitt pro Welle geschätzten Mittelwerte und die jeweiligen Interdezilbereiche sowie als Vergleichswerte die in der Shell-Jugendstudie ermittelten Mittelwerte der Indikatoren gibt Tabelle 9 wider.

**Tabelle 9:**

*Veränderungen der Mittelwerte (arithmetische Mittel -MW-) und der Streuungen (Interdezilbereiche -IDB-) von Welle zu Welle für die Gesamtheit der befragten Experten*

	1.Welle		2.Welle		3.Welle		Shell
	MW	IDB	MW	IDB	MW	IDB	MW
Harmonie	5.4	5.0-6.0	5.3	5.0-6.0	5.4	5.0-6.0	5.9
Soz. Macht	3.4	2.0-4.5	3.5	2.0-4.5	3.5	2.0-4.5	2.5
Freiheit	6.1	5.8-6.5	6.2	6.0-6.5	6.2	6.0-6.5	6.3
Soz. Ordnung	4.5	3.0-6.0	4.4	3.0-5.2	4.4	3.0-5.1	5.4
Anregung	5.5	4.0-6.3	5.7	5.0-6.1	5.7	5.0-6.0	5.7
Höflichkeit	3.6	2.0-5.0	3.5	2.0-4.5	3.4	2.0-4.0	5.3
Reichtum	4.4	3.0-5.5	4.6	4.0-5.5	4.6	4.0-5.0	4.4
Sicherheit	3.8	2.3-5.0	3.7	3.0-4.5	3.6	3.0-4.5	5.0
Kreativität	5.1	4.5-6.0	5.1	4.5-6.0	5.1	4.0-6.0	5.2
Frieden	6.0	5.0-7.0	6.1	5.0-7.0	6.0	5.0-6.5	6.6
Tradition	2.8	2.0-4.0	2.6	2.0-3.2	2.7	2.0-3.0	3.9
Loslösung	2.9	2.0-4.0	2.7	2.0-3.5	2.7	2.0-3.5	2.9
Familie	5.3	4.0-6.5	5.3	4.5-6.0	5.2	4.5-6.0	6.3
Natur	5.1	4.0-6.0	5.1	4.0-6.0	5.0	4.0-5.5	5.7
Abwechslung	5.2	4.0-6.0	5.3	4.0-6.0	5.2	4.8-6.0	5.8
Autorität	3.6	2.0-5.0	3.6	2.5-4.5	3.6	2.5-4.5	3.5
Freundschaft	5.8	5.0-7.0	5.7	5.0-6.1	5.8	5.0-6.0	6.5
Schönheit	3.8	3.0-5.0	3.6	3.0-4.0	3.8	3.0-4.0	5.2

In sechs Fällen<sup>9)</sup> können die in der dritten Welle von den Experten geschätzten Mittelwerte als besonders gute "Treffer" bezeichnet werden. Bei diesen Indikatoren unterscheiden sich die

9) In der Tabelle hervorgehoben.

tatsächlichen (in der Shell-Jugendstudie ermittelten) Mittelwerte von den geschätzten lediglich um höchstens 0.2 Skalenpunkte (bei der Schätzung der Wichtigkeit eines anregenden Lebens tritt sogar keine Differenz auf). Allerdings muß auch festgestellt werden, daß die Ergebnisse der dritten Welle im Vergleich zu denen der ersten Welle keine Verbesserung der Schätzung gebracht haben.

Auch bei drei weiteren Vorgaben liegen die wahren Werte innerhalb der Interdezilbereiche der letzten Schätzung, so daß bei insgesamt neun (von 18) Fällen zufriedenstellende Schätzwerte für die tatsächlichen Mittelwerte vorliegen. Zugleich muß aber auch von fünf Fehlschätzungen berichtet werden, in denen sich die Mittelwerte im Verlauf der Wellen nicht auf den wahren Wert zubewegt haben. Insbesondere bei den Wichtigkeitsurteilen für Höflichkeit/gute Umgangsformen, soziale Sicherheit/Stabilität der Gesellschaft und eine Welt der Schönheit/Schönheit der Natur und Künste weichen die finalen Schätzungen um 1,9 bzw. 1,4 und 1,4 Skalenpunkte vom tatsächlichen Mittelwert ab.

Insgesamt ist bemerkenswert, daß die Experten bei 14 von 18 Indikatoren den tatsächlichen Wert im Mittel unterschätzt haben, also allgemein dazu neigten, den Jugendlichen weniger stark ausgeprägte Wichtigkeitsurteile zuzuweisen, als sie tatsächlich vorhanden sind.

Korrelationsanalysen der Mittelwerte der von den Experten in den jeweiligen Wellen geschätzten Werte und den Mittelwerten der Shell Jugendstudie lieferten die folgenden Ergebnisse<sup>10)</sup>:

	B	R
1. Welle <-> Shell:	73.42%	0.8568
2. Welle <-> Shell:	67.60%	0.822
3. Welle <-> Shell:	67.44%	0.82

Die Korrelationskoeffizienten zwischen den geschätzten und den tatsächlichen Mittelwerten haben sich von Welle zu Welle zwar leicht verringert, zeigen aber insgesamt, daß zwischen geschätzten und tatsächlichen Werten ein relativ starker Zusammenhang besteht.

Die Streuung (betrachtet werden hier als Dispersionsmaß zunächst die Interdezilbereiche) um die geschätzten Werte hat sich von Welle zu Welle verringert. So betrug der Interdezilbereich in der ersten Welle durchschnittlich 2.15 Skalenpunkte, in der zweiten Welle 1.6 und schließlich in der dritten Welle noch 1.4 Skalenpunkte. Diese Verringerung der Streuung hatte allerdings zur Folge, daß zwar nach der ersten Welle noch 16 der in der Shell-Studie ermittelten Mittelwerte innerhalb des von den Experten geschätzten Interdezilbereiches lagen, nach der dritten Welle dagegen nur noch die bereits erwähnten neun.

Für die einzelnen Indikatoren veränderten sich die Variationskoeffizienten von Welle zu Welle wie in Tabelle 10 dargestellt.

---

10) Die Ergebnisse sind jeweils mit  $p < 0.001$  signifikant.

Tabelle 10:  
Variationskoeffizienten für die Mittelwertschätzungen  
(Standardabweichungen in Prozent der Mittelwerte)

	1. Welle	2. Welle	3. Welle	Shell
Harmonie	13.99	11.28	11.43	26.59
Soz. Macht	38.30	23.78	23.68	70.92
Freiheit	7.07	5.48	4.83	19.22
Soz. Ordnung	23.83	20.82	20.06	29.79
Anregung	16.30	8.86	8.42	25.23
Höflichkeit	32.86	23.92	22.62	31.67
Reichtum	23.52	14.06	12.90	43.79
Sicherheit	26.77	16.84	20.67	39.63
Kreativität	15.41	12.69	12.99	62.14
Frieden	15.28	11.76	9.65	15.47
Tradition	30.32	20.83	18.45	52.22
Loslösung	37.84	26.47	25.19	65.27
Familie	23.30	16.60	15.37	20.14
Natur	18.38	15.09	15.19	26.80
Abwechslung	19.18	15.39	13.02	24.50
Autorität	32.58	22.28	25.88	56.82
Freundschaft	12.78	9.56	7.09	15.70
Schönheit	26.25	18.40	12.18	34.00

Korrelationsanalysen zwischen den Variationskoeffizienten der Expertenschätzungen und denen der Shell-Studie ergaben, daß auch zwischen den Streuungen der beiden Item-Batterien ein signifikanter Zusammenhang besteht<sup>11)</sup>:

	B	R
1. Welle <-> Shell:	50.65%	0.71
2. Welle <-> Shell:	41.26%	0.64
3. Welle <-> Shell:	47.71%	0.69

Insgesamt sind die Ergebnisse der Mittelwertschätzungen aber differenzierter ausgefallen und zu bewerten als die in den vorigen Abschnitten beschriebenen Resultate der Verteilungsschätzungen. Als mögliche Ursache für die hier vermehrt aufgetretenen Fehlrteile vermuteten wir zunächst, daß sich die Experten bei ihrer Schätzung nicht die befragte Population der 13- bis 29-jährigen, sondern eine bestimmte Subpopulation von Jugendlichen - etwa Studenten in Westdeutschland - vorgestellt hatten, wie dies beispielsweise bei der Schätzung der Alternativfragen gezeigt werden konnte. Um diesem Problem nachzugehen, wurde in der ursprünglichen Datei der Shell-Jugendstudie systematisch nach solchen Untergruppen gesucht, in denen die Mittelwerte der Indikatoren zu den Wichtigkeitsurteilen den von den Experten geschätzten Ergebnissen

11) Die Ergebnisse sind jeweils mit  $p < 0.001$  signifikant.



entsprechen bzw. nahekommen. Dazu wurden die Variablen Alter (gruppiert), Berufsstatus und Ost-West-Filter benutzt. Die relativ große Stichprobe (N=4005) erlaubte es auch, verschiedene Kombinationen dieser Variablen zu überprüfen. Als Ergebnis stellte sich allerdings heraus, daß sich mit den genannten Variablen keine Subgruppen bilden lassen, auf die die von den Experten geschätzten (hier vor allem falschen) Werte zutreffen würden. Es ergab sich beispielsweise weder eine einer bestimmten Regel folgende Alters- oder Qualifikationsabhängigkeit, noch ein Bezug nur auf west- oder ostdeutsche Jugendliche. Damit liegt also keine systematische "Fehlorientierung" der Experten vor, sondern schlicht Unwissen.

Ein weiterer Analyseschritt ergab sich ebenfalls aus der Suche nach Gründen dafür, daß es den Experten durchaus gelungen war, die Hälfte der Mittelwerte der vorgegebenen Indikatoren relativ gut zu schätzen, andere dagegen weniger gut. Es wurde untersucht, ob die treffend geschätzten Indikatoren zu den Wichtigkeitsurteilen inhaltlich als zusammengehörig (z.B. zu bestimmten Wertetypen) bezeichnet werden können. Dazu wurde eine explorative Faktoranalyse<sup>12)</sup> über die Originaldaten der 18 in der Shell-Studie erfragten Wertorientierungen (N=4005) ausgeführt. Dabei wurden die folgenden Faktoren extrahiert:

Faktor	Eigenwert	Prozent der Varianz	Kum. Prozent
1	4.02431	22.4	22.4
2	1.94769	10.8	33.2
3	1.48038	8.2	41.4
4	1.30293	7.2	48.6

Die Faktorladungen lassen auf die folgende, in Tabelle 11 dargestellte, Zuordnung der Wertorientierungsindikatoren zu den Faktoren schließen<sup>13)</sup>.

Es zeigt sich, daß acht der neun "gut geschätzten" (also innerhalb des geschätzten Interdezilbereiches liegenden) Indikatoren aufgrund der Höhe ihrer Faktorladungen den Faktoren 2 bzw. 3 zuzuordnen sind. Nur der Indikator "Loslösung", dessen Mittelwert von den Experten ebenfalls relativ gut geschätzt worden war, erscheint als zu Faktor 4 gehörig. Dieses Ergebnis könnte darauf hindeuten, daß es unter den Wertorientierungen inhaltliche Komplexe (wie sie z.B. in einer Studie von Krebs 1992 mit "Macht" und "Stimulation" bezeichnet wurden) gibt, die von den Experten besser geschätzt werden können als andere. Um diese Hypothese zu verifizieren, sind weitere Analysen notwendig, die jedoch späteren Arbeiten vorbehalten bleiben müssen.

---

12) Faktorextraktion: Hauptkomponentenanalyse; Bestimmung der Faktorenzahl nach dem Kaiser-Kriterium (Eigenwert > 1); Rotation nach dem Varimax-Kriterium.

13) Die Zuordnung der Indikatoren zu den Faktoren erfolgte bei Faktorladungen > 0.5 (vgl. Konvention bei Backhaus et al. 1988, S.108).

Tabelle 11:  
Rotierte Faktor Matrix

	FAKTOR 1	FAKTOR 2	FAKTOR 3	FAKTOR 4
Harmonie*	.30545	.42900	-.11817	.04099
Soz. Macht*	-.07771	-.05863	<b>.72634</b>	.03335
Freiheit*	.25191	<b>.63316</b>	-.06706	-.08061
Soz. Ordnung	<b>.65855</b>	.20754	.16811	-.03406
Anregung*	.05022	<b>.72446</b>	.20264	.00375
Höflichkeit	<b>.64221</b>	.08717	.04002	.23095
Reichtum	.15393	.05648	<b>.67626</b>	-.10975
Sicherheit	<b>.67482</b>	-.03134	.31005	-.03301
Kreativität*	-.05277	<b>.53337</b>	.03376	.44336
Frieden	.47689	.32762	-.25506	.14524
Tradition	.35066	-.11841	.19519	<b>.57761</b>
Loslösung*	-.09411	-.01219	.07786	<b>.67690</b>
Familie	<b>.66042</b>	.13339	-.10480	.10387
Natur	.31842	.32038	-.22772	<b>.52790</b>
Abwechslung*	-.01356	<b>.66984</b>	.09889	.15520
Autorität*	.06854	.07175	<b>.69981</b>	.21760
Freundschaft	.36010	<b>.52480</b>	-.21419	.12650
Schönheit	.20957	.32002	-.04623	<b>.55329</b>

Die "gut geschätzten", also jeweils innerhalb des geschätzten Interdezilbereiches liegenden Indikatoren sind mit einem "\*" gekennzeichnet.  
Die Zuordnung der Indikatoren zu den Faktoren wird durch Fettdruck ausgewiesen.

Im folgenden sollen die Ergebnisse der Mittelwertschätzungen für die einzelnen Experten präsentiert werden. Zur besseren Übersicht wurde ein synthetisches Merkmal (die quadrierten Euklidischen Abstände der Schätzungen der Mittelwerte zu den jeweils wahren Werten pro Experte) gebildet. Damit wird zunächst eine Aussage darüber möglich, wie die einzelnen Probanden das Problem der Mittelwertschätzung insgesamt bewältigten.

Von den 20 Experten konnten 14 ihr Ergebnis der Mittelwertschätzungen zwischen der ersten und der dritten Welle insgesamt verbessern, drei Experten veränderten ihr Urteil über die Wellen nicht und bei drei Testteilnehmern war die Anfangsschätzung besser als die in der dritten Welle abgegebene. Bei fünf der Experten, deren Schätzung sich zwischen der ersten und der dritten Welle insgesamt verbesserte, lag die zweite abgegebene Schätzung näher am wahren Wert als die dritte (vgl. Tabelle 12).

Tabelle 12:  
*Quadrierte Euklidische Distanzen der von den Probanden geschätzten Mittelwerte zu den wahren Werten (Mittelwertfragen)*

Exp.nr.	Quadrierte Euklidische Abstände		
	1. Welle	2. Welle	3. Welle
01	37.52	23.72	28.72
02	54.72	47.92	47.92
03	17.32	15.07	26.57
04	12.04	12.04	12.04
05	13.32	8.72	13.12
06	<b>56.12</b>	<b>45.12</b>	<b>42.92</b>
07	16.47	11.27	12.62
08	33.92	12.23	25.62
09	<b>19.22</b>	<b>15.67</b>	<b>14.52</b>
10	17.92	18.68	18.22
11	19.91	14.66	16.65
12	62.77	62.77	54.57
13	14.53	13.90	15.10
14	21.72	21.72	21.72
15	<b>37.17</b>	<b>22.71</b>	<b>21.99</b>
16	21.62	21.62	21.62
17	<b>57.72</b>	<b>35.72</b>	<b>23.72</b>
18	32.72	22.32	22.32
19	<b>15.12</b>	<b>12.32</b>	<b>11.43</b>
20	<b>18.74</b>	<b>14.35</b>	<b>8.22</b>

Insbesondere bei den in der Tabelle 12 hervorgehobenen sechs Experten verlief der Annäherungsprozeß an die wahren Werte über die drei Wellen wie bei der Delphi-Methode günstigstenfalls zu erwarten.

Im folgenden soll nun überprüft werden, ob die gleichen Experten, die gut in der Lage waren, die Schätzungen zu den Alternativfragen abzugeben, auch die Schätzungen zu den Mittelwerten der Wichtigkeitsurteile gut absolvierten. Unsere Vermutung ist, daß dies nicht der Fall sein wird, da einerseits die Anforderungen an die geforderten Schätzleistungen jeweils sehr unterschiedlich sind und andererseits inhaltlich ein anderer Themenkomplex erfragt wird.

Zur Prüfung der Annahme wird zunächst die bei der Schätzung der Alternativfragen entstandene Klassifizierung in "bessere" und "schlechtere" Experten übernommen. Aufgrund modelltheoretischer Voraussetzungen<sup>14)</sup> kann nun nicht die diskriminatorische Bedeutung *aller* Mittelwertfragen für die Unterscheidung in die beiden Expertengruppen in einem Modell geprüft werden. Deshalb wird die oben ermittelte inhaltliche Zuordnung der einzelnen Indikatoren zu den Wertetypen bildenden Faktoren als Gruppierungskriterium angewendet.

14) Bei der Diskriminanzanalyse sollte der Umfang der Stichprobe wenigstens doppelt so groß sein wie die Anzahl der Merkmalsvariablen (vgl. Backhaus et al. 1988, S.213).

Als diskriminierende Variablen fungieren (gruppiert nach den vier Faktoren) die absoluten Beträge der Abweichungen der von den Experten in der dritten Welle abgegebenen Schätzungen der Mittelwerte für die Indikatoren von den wahren Werten. Im einzelnen ergaben sich die folgenden Modelle.

Faktor 1:

Einbezogene diskriminierende Variablen:

Absolute Beträge der Abweichungen der Expertenschätzungen (3. Welle) von den wahren Werten folgender Indikatoren:

- Wichtigkeit soziale Ordnung
- Wichtigkeit Höflichkeit
- Wichtigkeit Sicherheit
- Wichtigkeit Familie
- \* Zielvariable: Einteilung der Experten in zwei Gruppen

Die ermittelte Diskriminanzfunktion für den ersten Faktor ist mit einem  $\chi^2$ -Wert von 4.843 bei vier Freiheitsgraden nicht signifikant. Sieben der zwanzig Experten wurden aufgrund der ermittelten Diskriminanzfunktion der jeweils falschen Gruppe zugeordnet.

Faktor 2:

Einbezogene diskriminierende Variablen:

Absolute Beträge der Abweichungen der Expertenschätzungen (3. Welle) von den wahren Werten folgender Indikatoren:

- Wichtigkeit Freiheit
- Wichtigkeit Anregung
- Wichtigkeit Kreativität
- Wichtigkeit Abwechslung
- Wichtigkeit Freundschaft
- \* Zielvariable: Einteilung der Experten in zwei Gruppen

Die ermittelte Diskriminanzfunktion für den zweiten Faktor ist mit einem  $\chi^2$ -Wert von 1.531 bei fünf Freiheitsgraden nicht signifikant. Ebenfalls sieben der zwanzig Experten wurden aufgrund der ermittelten Diskriminanzfunktion der jeweils falschen Gruppe zugeordnet.

Faktor 3:

Einbezogene diskriminierende Variablen:

Absolute Beträge der Abweichungen der Expertenschätzungen (3. Welle) von den wahren Werten folgender Indikatoren:

- Wichtigkeit soziale Macht
- Wichtigkeit Reichtum
- Wichtigkeit Autorität
- \* Zielvariable: Einteilung der Experten in zwei Gruppen

Die ermittelte Diskriminanzfunktion für den dritten Faktor ist mit einem  $\chi^2$ -Wert von 6.206 bei drei Freiheitsgraden nicht signifikant. Bei diesem Faktor wurden sechs der zwanzig Experten aufgrund der ermittelten Diskriminanzfunktion der jeweils falschen Gruppe zugeordnet.

Faktor 4:

Einbezogene diskriminierende Variablen:

Absolute Beträge der Abweichungen der Expertenschätzungen (3. Welle) von den wahren Werten folgender Indikatoren:

- Wichtigkeit Tradition
- Wichtigkeit Loslösung
- Wichtigkeit Natur
- Wichtigkeit Schönheit
- \* Zielvariable: Einteilung der Experten in zwei Gruppen

Auch die ermittelte Diskriminanzfunktion für den vierten Faktor ist mit einem  $\chi^2$ -Wert von 4.854 bei vier Freiheitsgraden nicht signifikant. Vier der zwanzig Experten wurden aufgrund der ermittelten Diskriminanzfunktion der jeweils falschen Gruppe zugeordnet.

Da die Zuordnungsfehler bei den einzelnen Diskriminanzfunktionen jeweils unterschiedliche Experten betrafen, ist auch durch eine leichte Korrektur der Expertenklassifizierung kein anderes Ergebnis zu erwarten.

Das gleiche Resultat erbrachte der Versuch, die Expertenklassifizierung aus den Aufgaben zur Schätzung von Antwortalternativen auf die Verteilungsschätzungen (vgl. Abschnitt 6.2.) zu übertragen. Als diskriminierende Variablen wurden hier die  $\chi^2$ -Prüfgrößen eingesetzt, die jeweils für die einzelnen Expertenschätzungen der dritten Welle für die fünf Fragen zu Indikatoren mit vierstufigem Antwortmuster ermittelt worden waren. Die so berechnete Diskriminanzfunktion ist mit  $p=0.508$  nicht signifikant, nur 68% der Experten wurden aufgrund der Funktion richtig in die ursprünglichen beiden Expertengruppen eingeordnet.

## Fazit

Die Übernahme der bei den Alternativfragen ermittelten Expertengruppierung in erfolgreichere und weniger erfolgreiche Schätzer scheint sowohl für die Mittelwert- als auch die Verteilungsfragen nicht gerechtfertigt zu sein<sup>15)</sup>. Demnach gibt es keine a priori für Delphi-Schätzungen besonders geeigneten Experten, sondern die Fähigkeit, Antworten schätzen zu können, scheint abhängig zu sein von den jeweils spezifischen Erfahrungen. So kann vermutet werden, daß die jeweils unterschiedlichen Erfahrungen der ZUMA-Experten, z.B. beim Umgang mit Jugendlichen, bei der Beschäftigung mit den inhaltlichen Problemen, bei der Beurteilung statistischer Verteilungen usw. jeweils einen spezifischen Beitrag für das Gelingen der Schätzungen geleistet haben. Dies spricht letztlich auch für die Validierung des gesamten Delphi-Ansatzes. So nutzen die Experten tatsächlich ihr jeweils spezifisches Wissen, um die ihnen vorgelegte Aufgabenstellung zu lösen.

Weiterhin kann aufgrund dieses Ergebnisses auch vermutet werden, daß es nur wenige von uns am ZUMA rekrutierte Experten gibt, die absolut *nicht* dazu in der Lage sind, auch nur bei einer Aufgabenart die verlangten Schätzungen erfolgreich vorzunehmen. Dies belegten gerade die letzten Auswertungsschritte. Unterstrichen werden soll nochmals die Bedeutung, die einer sorgfältigen Expertenrekrutierung prinzipiell beigemessen werden muß.

Nicht die erwartete Bedeutung hatten dagegen die den Originaldaten zugrunde liegenden Verteilungsformen bzw. (bei den Mittelwertschätzungen) die Streuungen. Aufgrund dieses Tests kann vermutet werden, daß der Erfolg der Schätzungen eher auf die Kompetenz der Experten als auf bestimmte Antwortverteilungen zurückzuführen ist. Dieses Ergebnis spricht für weitere Tests in diese Richtung.

## 6.5. Auswertung der ergänzenden Expertenbefragung (vierte Welle)

Zunächst werden kurz die Ergebnisse der Zusatzbefragung (vgl. Anhang 4) im Überblick referiert:

Trotz der methodischen Erfahrungen der ZUMA-Mitarbeiter schätzten vier Probanden das Anliegen des Tests (1. Frage) auf einer 7-stufigen Skala als für sie eher nicht verständlich ein (Antwortstufe 5 und 6). Wir interpretieren dies zunächst als einen Hinweis auf die Kompliziertheit des Anliegens dieser Art von Delphi-Tests.

Während sich nur drei Experten bisher intensiver mit Fragen des Zeitbewußtseins (2. Frage) befaßt haben<sup>16)</sup>, trifft dies bei den Wichtigkeitsurteilen (3. Frage) auf neun Probanden zu. Damit gehörten der Befragtengruppe offenbar nur wenige Personen an, die sich bereits explizit inhaltlich mit den zu schätzenden Sachverhalten auseinandergesetzt haben. Die (relativ "gleichmäßige")

---

15) Hierzu sind ergänzende (insbesondere konfirmatorische) Analysen vorgesehen.

16) Die ursprünglich bei der Befragung benutzten 7-stufigen Skalen werden hier stark vereinfacht ausgewertet. So fassen wir jeweils die drei extremen Antwortkategorien auf beiden Seiten zusammen.

Antwortverteilung auf die dritte Frage erlaubt es immerhin jedoch, die Expertengruppe entsprechend weiter zu gruppieren.

Die Frage nach dem Zustandekommen der ersten Schätzung läßt erkennen, daß hier - wie bereits vermutet - vor allem die eigenen Erfahrungen aus dem Umgang mit Jugendlichen eine Rolle spielten und daß außerdem spontan nach dem Gefühl geantwortet wurde. Am wenigsten wurden - angesichts der Antworten auf die Fragen 2 und 3 völlig verständlich - für die ersten Schätzungen auf eigene Erfahrungen mit ähnlichen Fragestellungen (Fragen 4a) bis d)) zurückgegriffen.

Bei der Frage nach den Schätzungen der folgenden Wellen wurde am häufigsten angegeben, daß hier über die rückgemeldete statistische Verteilungsform nachgedacht worden war (Fragen 5a) bis d)). Auch der Abstand der eigenen ursprünglichen Schätzung vom Gruppenmittel sowie ein nochmaliges Nachdenken über den Inhalt spielten noch bei jeweils 13 Experten eine Rolle.

Acht Befragte halten die am ZUMA rekrutierten Experten für eher kompetent, um mit Hilfe einer solchen Delphi-Studie Befragungsergebnisse zu schätzen (Frage 6). Daraus läßt sich ein Hinweis auf die (jeweils) den rückgemeldeten Gruppenergebnissen beigemessene Bedeutung ableiten.

Diese müßte bei jenen Befragten deutlich höher sein als bei den übrigen.

Eher viele Kontakte mit Jugendlichen gaben 14 Experten an (Frage 7), während nur fünf Befragte eine Antwortmöglichkeit wählten, die in die entgegengesetzte Richtung zielte.

Als Studienrichtung dominierte die Soziologie (11 Experten, 8. Frage).

Die immense Erfahrung beim Umgang mit empirisch-sozialwissenschaftlichen Problemen wird in der folgenden 9. Frage deutlich. So haben sich die am ZUMA befragten Experten im Durchschnitt seit 11 Jahren mit solchen Problemen befaßt.

Von Welle zu Welle zurückgegangen ist die für das Ausfüllen benötigte Zeit (Frage 10). Für die erste Welle wurden noch durchschnittlich 17, für die zweite 9 und für die dritte nur noch 6 Minuten benötigt.

Da, wie bereits berichtet wurde, eine Klassifizierung der Experten aufgrund ihres Schätzerfolges nicht ohne weiteres möglich ist, ist es schwierig, mit diesen Daten weitere Auswertungsschritte vorzunehmen.

## **7. Zusammenfassung und Ausblick**

Dieser erste Test zur Validierung der Ergebnisse einer Delphi-Befragung am ZUMA hat umfangreiches empirisches Material erbracht und war allein schon deshalb aus erhebungsmethodischer Sicht wertvoll. Vor allem aber hat er mehrheitlich gelungene Schätzungen, die sich in Richtung auf die wahren Werte bewegen, erzeugt. So kann festgestellt werden:

1. Es ist generell nach drei Befragungswellen besser möglich, Aussagen über die Antwortverteilungen bei einer repräsentativen Befragung unter Jugendlichen zu treffen als bei einem nur einmaligen Versuch.



2. Die zusammengefaßten Gruppenergebnisse sind von höherer Qualität als nur isolierte Schätzungen einzelner Experten.

Damit sind zunächst zwei wichtige empirische Hinweise für das Funktionieren dieser Vorgehensweise gewonnen worden.

3. Es ließ sich demonstrieren, daß die Form der Antwortverteilungen in den ursprünglichen Daten keinen generellen Einfluß auf den Erfolg bzw. die Fehler solcher Schätzungen hat. Damit liegt ein weiteres Argument vor, um die Nutzung der Delphi-Technik für die Schätzung von Ergebnissen allgemeiner Bevölkerungsbefragungen mit Erfolg weiter zu bearbeiten.

4. Auch das in Abschnitt 3. dargestellte methodologische Untersuchungsprogramm hat sich damit als tragfähig erwiesen. Die hier formulierten Ziele - die Delphi-Technik zu validieren und dabei Schätzungen von Ergebnissen aus Bevölkerungsbefragungen zu benutzen - sind sinnvoll und versprechen, zu weiteren interessanten Ergebnissen zu führen.

Außerdem wurde auf diese Weise eine eigene Verbindung zwischen qualitativer und quantitativer Datengewinnung vorgenommen. Die hier gewonnenen Ergebnisse bieten die Möglichkeit für eigene methodologische Analysen.

5. Über die individuellen Fähigkeiten der Experten, bei Delphi-Studien verschiedene Typen von Schätzungen über Antwortverteilungen bei Bevölkerungsbefragungen vorzunehmen, konnten ebenfalls erste Aussagen gewonnen werden. Wichtig erscheint vor allem, daß sich bei diesem Test empirisch keine (Unter-)Gruppe extrahieren ließ, die besonders dazu prädestiniert wäre, solche Schätzungen vorzunehmen. Offenbar richtet sich das Vermögen der Experten, richtige Schätzurteile abzugeben, nach den jeweiligen konkreten Besonderheiten der einzelnen Aufgabentypen.

Dieser erste Test konnte und wollte die umfangreichen methodischen Probleme von Delphi-Erhebungen nicht erschöpfend bearbeiten. So stellen sich nun mehr und zugleich differenziertere Fragen als vor Beginn dieses Tests. Verschiedene Aspekte erfordern eine genauere Betrachtung:

- Die gewonnenen Hinweise vermitteln erste, oftmals noch schwer zu systematisierende Eindrücke vom Umgang mit diesem Verfahren. Eine spezielle, das Funktionieren dieser Methode beschreibende Theorie ist noch nicht in Sicht. Dies gilt ebenfalls für deren Verbindung mit den anderen erhebungsmethodischen Verfahren zur Datengewinnung.

Auch Reuband (1990, S.721) sieht bei seiner Analyse zu Interviewfälschungen "substantielle Implikationen für die soziologische Theoriebildung", die sich der Frage zuzuwenden hat, wie es möglich ist, daß "sich Menschen Vorstellungen über Personen in anderen sozialen Lagen machen". Er sieht vor allem in der Alltagskommunikation eine Erklärungsvariable für dieses Phänomen.

So versteht sich auch dieser Test als erster Schritt zu einer Theorie der Delphi-Technik.

- Die benutzte Methodik wurde in diesem ersten Test nicht gezielt variiert. Damit können keine Aussagen darüber gemacht werden, welchen Einfluß beispielsweise eine andere Gestaltung der



Rückinformation (Nutzung eines anderen Streuungsmaßes, Verzicht auf die Rückmeldungen der eigenen vorangegangenen Schätzung usw.) haben würde.

- Noch nicht ausreichend beantwortet werden konnte die Frage nach den Bestimmungsgrößen für den Erfolg dieser Schätzungen. Eine systematische Zusammenstellung der Determinanten für den Erfolg der Schätzungen kann jedoch - auch aufgrund der relativ geringen Zahl der beteiligten Experten sowie allein aufgrund dieses ersten Tests - nicht geleistet werden.

Mitunter sind wir auf Hinweise gestoßen, die als mögliche Erklärung für die begangenen Schätzfehler Plausibilität besitzen.

Daraus lassen sich ergänzend drei weitere Hypothesen ableiten:

1. Die Experten lassen sich - mehr oder weniger unbewußt - bei ihren Schätzungen von konkreten eigenen Erfahrungen leiten. Z.B. basierten die Vorstellungen der westdeutschen Experten über Jugendliche in Deutschland auf ihren Erfahrungen mit vor allem westdeutschen Jugendlichen. Damit sind teilweise unbefriedigende Schätzungen bei von einer größeren Subpopulation - z.B. den ostdeutschen Jugendlichen - signifikant anders beantworteten Fragen zu erwarten gewesen.

2. Die Experten können die in den Daten latent verborgenen Strukturen nicht mit gleicher Sicherheit schätzen.

Es kommt zu (nichtzufällig) unterschiedlichen Schätzerfolgen bei relativ umfangreichen Indikatorenbatterien, die komplexen Sachverhalten gewidmet sind - in unserem Test beispielsweise bei Wichtigkeitsfragen. Bestimmte mit den Urdaten faktoranalytisch ermittelbare Dimensionen werden mit jeweils unterschiedlichem Erfolg geschätzt.

3. Es besteht eine Tendenz, wonach kompliziertere Aufgabenstellungen (z.B. die Schätzungen von Antwortverteilungen mit unterschiedlichen Stufen, die anspruchsvollere Überlegungen erfordern) besser geschätzt werden können als vergleichsweise einfach erscheinende Angaben (z.B. Mittelwertschätzungen).

## Literatur

- Abrams, M. (1949), Possibilities and problems of Group Interviewing, in: Public Opinion Quarterly, Bd. 13.
- Albach, H. (1970): Informationsgewinnung durch strukturierte Gruppenbefragung. Die Delphi-Methode. In: Zeitschrift für Betriebswirtschaft 40/1970 (Ergänzungsheft): 11-26.
- Alemann, H.v. (1977): Der Forschungsprozeß. Eine Einführung in die Praxis der empirischen Sozialforschung. Stuttgart: Teubner.
- Atteslander, P. (1984), Methoden der empirischen Sozialforschung, Berlin, New York: W. de Gruyter.
- Backhaus, K., Erichson, B., Plinke, W., Schuchard-Fischer, Chr., Weiber, R. (1988), Multivariate Analysemethoden. Eine Anwendungsorientierte Einführung, Berlin Heidelberg New York London Paris Tokyo: Springer
- Barsalou, L. W. (1987), The instability of graded structure: Implications for the nature of concepts, in: U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization, Cambridge: Cambridge University Press, S. 101 - 140.
- Barsalou, L.W. (1989), Intraconcept similarity and its implications for interconcept similarity, in: S. Vosniadou/ A. Ortony (Hrsg.), Similarity and analogical reasoning, Cambridge: Cambridge University Press, S. 76-121.
- Bodenhausen, G. V., Wyer, R. S. (1987), Social cognition and social reality: Information acquisition and use in the laboratory and the real world, in: H.-J. Hippler, N. Schwarz, S. Sudman (Eds.) Social information processing and survey methodology, New York: Springer, S. 6 - 41.
- Bortz, J. (1984), Lehrbuch der empirischen Forschung für Sozialwissenschaftler, Berlin Heidelberg New York Tokio: Springer.
- Breed, W., Ktsans, Th. (1961), Pluralistic Ignorance in the Process of Opinion Formation, in: Public Opinion Quarterly, Bd. 25.
- Bronner, R. et al. (1991): Anforderungen an Spitzen-Führungskräfte: Ergebnisse einer Delphi-Studie. In: Zeitschrift für Betriebswirtschaft Nr.11: 1227-1242.
- Esser, H. (1986), Können Befragte lügen? Zum Konzept des 'wahren Wertes' im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 38, S. 314-336.
- Friedrichs, J. (1980), Methoden empirischer Sozialforschung, Opladen: Westdeutscher Verlag.
- Gewald, K. (1972): Die Delphi-Methode als Instrument technischer Prognosen - Praktische Erfahrungen. S. 13-18. In: Blohm, H./ Steinbuch, K., Technische Prognosen in der Praxis. Methoden, Beispiele, Probleme. Düsseldorf.
- Häder, M., Häder, S. (1994), Die Grundlagen der Delphi-Methode. ZUMA-Arbeitsbericht 94/02.
- Hartmann, H. (1979): Empirische Sozialforschung. München.

Higgins, E. T., King, G. (1981), Accessibility of social constructs: Information processing consequences of individual and contextual variability, in: N. Cantor, J. F. Kihlstrom (Eds.), Personality, cognition, and social interaction. Hillsdale: Erlbaum.

König, R. (1974): Handbuch der empirischen Sozialforschung. Bände 2, 3a, 3b. Stuttgart: Enke.

Köhler, G. (1992): Methodik und Problematik einer mehrstufigen Expertenbefragung. In: Jürgen H. P. Hoffmeyer-Zlotnik (Hrsg.), Analyse verbaler Daten. Über den Umgang mit qualitativen Daten. Opladen: Westdeutscher Verlag.

Kohr, H.-U. (1992), Zeit-, Lebens- und Zukunftsorientierungen, in: Jugend '92. Lebenslagen, Orientierungen und Entwicklungsperspektiven im vereinten Deutschland, hrsg. v. Jugendwerk der Deutschen Shell, Leske + Budrich: Opladen.

Krebs, D. (1992), Werte in den alten und neuen Bundesländern, in: Jugend '92. Lebenslagen, Orientierungen und Entwicklungsperspektiven im vereinten Deutschland, hrsg. v. Jugendwerk der Deutschen Shell, Leske + Budrich: Opladen.

Kromrey, H. (1980): Empirische Sozialforschung. Modelle und Methoden der Datenerhebung und Datenauswertung. Opladen: Leske & Budrich.

Linestone, H.A./ Turoff, M. (Hrsg.) (1975): The Delphi Method. London u.a.: Addison-Wesley.

Lohse, H., Ludwig, R. (1982), Prüfstatistik - Ein programmierter Lehrgang, Leipzig.

Mangold, W. (1967), Gruppendiskussion, in: Handbuch der empirischen Sozialforschung, Hrsg. R. König, Bd. 2, Grundlegende Methoden und Techniken der empirischen Sozialforschung, erster Teil, Ferdinand Enke: Stuttgart, S. 228 - 256.

Mayntz, R./ Holm, K./ Hübner, P. (1978): Einführung in die Methoden der empirischen Soziologie. Opladen.

Nehnevajsa, J. (1967), Analyse von Panel-Befragungen, in: Handbuch der empirischen Sozialforschung, Hrsg. R. König, Bd. 2, Grundlegende Methoden und Techniken der empirischen Sozialforschung, erster Teil, Ferdinand Enke: Stuttgart, S. 191 - 227.

Overbury, R.E. (1969): Technological Forecasting. A Criticism of the Delphi-Technique. In: Long-Range-Planning, vol. 1, No 4: 76-77.

Pollock, F. (1955), Hrsg., Gruppenexperiment - ein Studienbericht, Frankfurt (Main).

Reuband, K.-H. (1990), Interviews, die keine sind. "Erfolge" und "Mißerfolge" beim Fälschen von Interviews, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, Jg. 42, Heft 4, S. 706 - 733.

Roth, E. (Hrsg.) (1984): Sozialwissenschaftliche Methoden. München/ Wien.

Rowlands, D.G. (1969): Technological Forecasting and the Delphi-Technique: A Reply. In: Long-Range-Planning, Vol. 2, No. 2: 78-79.

Saliger, E./ Kunz, C. (1981): Zum Nachweis der Effizienz der Delphi-Methode. In: Zeitschrift für Betriebswirtschaft 51/ 1981: 470-480.

Schnell, R./ Hill, P. B./ Esser, E. (1992): Methoden der empirischen Sozialforschung. München/ Wien.

Schwarz, N. (1991), Assimilation und Kontrast in der Urteilsbildung: Implikationen für Fragereihenfolgeeffekte, in: ZUMA-NACHRICHTEN 29, S. 70 - 86.

Schwarz, N., Strack, F., Hippler, H.-J. (1991), Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebiets, in: Psychologische Rundschau, 42. Jg., S. 175 - 186.

Sherif, M. (1937, The Psychology of Social Norms, New York.

Smith, G. H. (1955), Warum Kunden kaufen, München.

Storm, R., (1979), Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle, Leipzig.

Strack, F. (1991), Kognitive und kommunikative Einflüsse in standardisierten Befragungssituationen, Habil-Arbeit.

Thurestone, L.L., Chave, E.J. (1929), The measurement of attitudes. Chicago, University of Chicago Press.

## Anlagenverzeichnis

1. Fragebogen für die Erhebung der ersten Welle	51
2. Fragebogen für die Erhebung der zweiten Welle	56
3. Fragebogen für die Erhebung der dritten Welle	61
4. Fragebogen für die Erhebung der vierten Welle (Zusatzbefragung)	66

## Anlage 1

Januar 1994

Im Rahmen unserer methodischen Tests zur Delphi-Technik wollen wir uns zuerst an Kollegen vom ZUMA wenden. Wir meinen, daß ihr in bezug auf Umfragen als Experten zu bezeichnen seid. Wir möchten mit Euch den Versuch wagen, in mehreren Wellen Antworten einer Bevölkerungsbefragung zu schätzen. Es versteht sich von selbst, daß das Ganze nur dann sinnvoll ist, wenn niemand schummelt und sich die Originaldaten besorgt.

Wir wenden uns zunächst nur an einen Teil der ZUMA-Mitarbeiter. Dies ist für unser heutiges Anliegen ausreichend. Gleichzeitig behalten wir uns damit die Möglichkeit offen, mit den anderen Kollegen zu einem späteren Zeitpunkt diesen Test zu wiederholen.

Im Jahr 1992 wurden in einer repräsentativen Befragung 4005 Jugendliche in Deutschland im Alter von 13 bis 29 Jahren nach ihrer Meinung zu verschiedenen Sachverhalten befragt (Shell-Jugendstudie '92). Einige der Fragen sind im folgenden aufgeführt.

Wir bitten Euch nun zu schätzen, wie die Antworten bei einzelnen Fragen in dieser Untersuchung ausgefallen sind. Uns interessieren sowohl Antwortverteilungen als auch Mittelwerte. Gebt also bitte jeweils den gesuchten Wert an.

**Berücksichtigt dabei unbedingt: Es interessiert hier nicht Eure persönliche Ansicht zu den Sachverhalten!**

Geht bitte davon aus, daß keine Antwortverweigerungen ("Missings") aufgetreten sind.

Wie bei Delphi-Studien üblich, möchten wir Euch gezielte Rückinformationen über die Antworten des Expertengremiums in den einzelnen Wellen geben. Um den Verlauf der Meinungsbildung detailliert verfolgen zu können, benötigen wir Euren Namen auf dem Fragebogen, d.h. die Daten werden auch personenbezogen gespeichert. Es ist selbstverständlich, daß wir die Angaben vertraulich behandeln und daß die Teilnahme freiwillig ist.

Bitte legt die ausgefüllten Fragebögen bis nächsten Mittwoch, den 19. Januar, zurück in unser Fach.

Vielen Dank für Eure Mitarbeit!

Michael und Sabine Häder

## 1. Welle

### Verteilungsschätzungen:

#### 1. Frage (24.1)<sup>17)</sup>

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

Heute ist heute und morgen ist morgen.

überhaupt nicht	....%
weniger gut	....%
gut	....%
sehr gut	....%

#### 2. Frage (24.2.) (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht	....%
weniger gut	....%
gut	....%
sehr gut	....%

#### 3. Frage (24.6.) (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht	....%
weniger gut	....%
gut	....%
sehr gut	....%

#### 4. Frage (24.10.) (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht	....%
weniger gut	....%
gut	....%
sehr gut	....%

#### 5. Frage (24. 23.) (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht	....%
weniger gut	....%
gut	....%
sehr gut	....%

---

17) Die Zahl in den Klammern gibt jeweils an, an welcher Stelle die Frage im Originalfragebogen stand.

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb zusammen 100% ergeben.

**6. Frage (25.a.)**

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

....%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

....%

**7. Frage (25.b.) (gleicher Fragetext)**

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

....%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

....%

**8. Frage (25. c.) (gleicher Fragetext)**

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

....%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

....%

**9. Frage (25. d.) (gleicher Fragetext)**

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

....%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

....%



## Mittelwertschätzungen

### 10. Frage (27.)

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

01. Innere Harmonie (in Frieden mit mir selbst)

Mittelwert: ...

02. Soziale Macht (Kontrolle über andere, Dominanz)

Mittelwert: ...

03. Freiheit (Freiheit des Handelns und des Denkens)

Mittelwert: ...

04. Soziale Ordnung (Stabilität der Gesellschaft)

Mittelwert: ...

05. Ein anregendes Leben (anregende Erfahrungen)

Mittelwert: ...

06. Höflichkeit (gute Umgangsformen)

Mittelwert: ...

07. Reichtum (materieller Besitz, Geld)

Mittelwert: ...

08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)

Mittelwert: ...

09. Kreativität (Originalität, Phantasie)

Mittelwert: ...

10. Eine Welt in Frieden (frei von Krieg und Konflikt)

Mittelwert: ...

11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)

Mittelwert: ...

12. Loslösung (von weltlichen Belangen)

Mittelwert: ...

13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)

Mittelwert: ...

14. Einheit mit der Natur (Einpassung in die Natur)

Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

Mittelwert: ...

Nochmals vielen Dank für Eure Mitarbeit, wir melden uns in Kürze wieder.

## Anlage 2

Januar 1994

Vielen Dank für Eure Antworten! Unser Delphi-Test geht nun in die nächste Runde. Wir teilen Euch dazu die Ergebnisse der ersten Welle mit und bitten um ein zweites Urteil. Zur Erinnerung legen wir Euch Eure Antworten vom letzten Mal bei.

Bitte denkt wieder an folgendes:

- Schätzt bitte - wie schon beim letzten Mal - die Ergebnisse einer Befragung, bei der 4005 Jugendliche in Deutschland im Alter von 13 bis 29 Jahren um ihrer Meinung zu verschiedenen Sachverhalten gebeten wurden (Shell-Jugendstudie 1992).
- Geht bitte wieder davon aus, daß keine Antwortverweigerungen ("Missings") aufgetreten sind.
- Nicht vergessen: Es interessiert hier nicht Eure persönliche Ansicht zu den Sachverhalten!
- Wir geben Euch als Rückinformation an: die Mittelwerte (MW), die Anzahl der gültigen Fälle (N) und als Streuungsmaß den Interdezilbereich (IDB), d.h. die Streubreite der mittleren 80 Prozent aller Fälle.

Bitte legt die ausgefüllten Fragebögen möglichst bis nächsten Freitag, den 4. Februar, zurück in unser Fach. Danach melden wir uns wieder.

Vielen Dank für Eure Mitarbeit!

Michael und Sabine Häder

## 2. Welle

### Verteilungsschätzungen:

#### 1. Frage (24.1)<sup>18)</sup>

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

Heute ist heute und morgen ist morgen.	MW	IDB	N	
überhaupt nicht	20%	10 - 30%	19	....%
weniger gut	31%	20 - 40%	19	....%
gut	30%	13 - 50%	19	....%
sehr gut	19%	5 - 25%	19	....%

#### 2. Frage (24.2.) (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht	15%	5 - 20%	19	....%
weniger gut	34%	20 - 50%	19	....%
gut	33%	20 - 50%	19	....%
sehr gut	18%	10 - 20%	19	....%

#### 3. Frage (24.6.) (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht	10%	0 - 18%	19	....%
weniger gut	21%	10 - 30%	19	....%
gut	40%	20 - 60%	19	....%
sehr gut	29%	15 - 50%	19	....%

#### 4. Frage (24.10.) (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht	29%	15 - 40%	19	....%
weniger gut	41%	30 - 50%	19	....%
gut	20%	10 - 30%	19	....%
sehr gut	10%	5 - 20%	19	....%

#### 5. Frage (24. 23.) (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht	26%	10 - 40%	19	....%
weniger gut	35%	20 - 45%	19	....%
gut	24%	15 - 40%	19	....%
sehr gut	15%	3 - 20%	19	....%

---

18) Die Zahl in den Klammern gibt jeweils an, an welcher Stelle die Frage im Originalfragebogen stand.

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb zusammen 100% ergeben.

#### 6. Frage (25.a.)

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

MW: 41% IDB: 30 - 60% N: 20 ....%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

MW: 59% IDB: 40 - 70% N: 20 ....%

#### 7. Frage (25.b.) (gleicher Fragetext)

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

MW: 49% IDB: 30 - 70% N: 19 ....%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

MW: 51% IDB: 30 - 70% N: 19 ....%

#### 8. Frage (25. c.) (gleicher Fragetext)

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

MW: 63% IDB: 30 - 80% N: 20 ....%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

MW: 37% IDB: 20 - 70% N: 20 ....%

#### 9. Frage (25. d.) (gleicher Fragetext)

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

MW: 50% IDB: 30 - 70% N: 20 ....%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

MW: 50% IDB: 30 - 70% N: 20 ....%

## Mittelwertschätzungen

### 10. Frage (27.)

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

#### 01. Innere Harmonie (in Frieden mit mir selbst)

MW: 5,4      IDB: 5,0 - 6,0      N: 20      geschätzter Mittelwert: ...

#### 02. Soziale Macht (Kontrolle über andere, Dominanz)

MW: 3,4      IDB: 2,0 - 4,5      N: 20      geschätzter Mittelwert: ...

#### 03. Freiheit (Freiheit des Handelns und des Denkens)

MW: 6,1      IDB: 5,8 - 6,5      N: 20      geschätzter Mittelwert: ...

#### 04. Soziale Ordnung (Stabilität der Gesellschaft)

MW: 4,5      IDB: 3,0 - 6,0      N: 20      geschätzter Mittelwert: ...

#### 05. Ein anregendes Leben (anregende Erfahrungen)

MW: 5,5      IDB: 4,0 - 6,3      N: 20      geschätzter Mittelwert: ...

#### 06. Höflichkeit (gute Umgangsformen)

MW: 3,6      IDB: 2,0 - 5,0      N: 20      geschätzter Mittelwert: ...

#### 07. Reichtum (materieller Besitz, Geld)

MW: 4,4      IDB: 3,0 - 5,5      N: 19      geschätzter Mittelwert: ...

#### 08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)

MW: 3,8      IDB: 2,3 - 5,0      N: 20      geschätzter Mittelwert: ...

#### 09. Kreativität (Originalität, Phantasie)

MW: 5,1      IDB: 4,5 - 6,0      N: 20      geschätzter Mittelwert: ...

#### 10. Eine Welt in Frieden (frei von Krieg und Konflikt)

MW: 6,0      IDB: 5,0 - 7,0      N: 20      geschätzter Mittelwert: ...

11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)

MW: 2,8      IDB: 2,0 - 4,0      N: 20      geschätzter Mittelwert: ...

12. Loslösung (von weltlichen Belangen)

MW: 2,9      IDB: 2,0 - 4,0      N: 20      geschätzter Mittelwert: ...

13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)

MW: 5,3      IDB: 4,0 - 6,5      N: 20      geschätzter Mittelwert: ...

14. Einheit mit der Natur (Einpassung in die Natur)

MW: 5,1      IDB: 4,0 - 6,0      N: 20      geschätzter Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

MW: 5,2      IDB: 4,0 - 6,0      N: 20      geschätzter Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

MW: 3,6      IDB: 2,0 - 5,0      N: 20      geschätzter Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

MW: 5,8      IDB: 5,0 - 7,0      N: 20      geschätzter Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

MW: 3,8      IDB: 3,0 - 5,0      N: 20      geschätzter Mittelwert: ...

Jetzt geht es endlich auch einmal um Eure Meinung! Zusätzlich bitten wir Euch nämlich die folgende Frage zu beantworten:

Wie sicher wart ihr Euch bei der Beantwortung der Fragen? Wenn ihr Euch völlig unsicher wart, kreuzt bitte das Kästchen ganz links an, wenn ihr Euch völlig sicher wart, kreuzt das Kästchen ganz rechts an. Mit den Kästchen zwischen den beiden Endpunkten könnt Ihr abstufen.

völlig  
unsicher

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

völlig  
sicher

Nochmals vielen Dank für Eure Mitarbeit.

Nochmals vielen Dank für Eure Antworten! Wir gehen nun mit unserem Delphi-Test in die dritte und wahrscheinlich letzte Runde. Wir teilen Euch dazu wieder die Ergebnisse der vorangegangenen Welle mit und bitten um ein weiteres Urteil. Zur Erinnerung legen wir Euch auch wieder Eure Antworten vom letzten Mal bei.

Bitte denkt nochmals an folgendes:

- Schätzt bitte - wie schon bei den beiden letzten Malen - die Ergebnisse einer Befragung, bei der 4005 Jugendliche in Deutschland im Alter von 13 bis 29 Jahren um ihrer Meinung zu verschiedenen Sachverhalten gebeten wurden (Shell-Jugendstudie 1992).
- Geht bitte wieder davon aus, daß keine Antwortverweigerungen ("Missings") aufgetreten sind.
- Nicht vergessen: Es interessiert hier nicht Eure persönliche Ansicht zu den Sachverhalten!
- Wir geben Euch als Rückinformation wieder an: die Mittelwerte (MW), die Anzahl der gültigen Fälle (N) und als Streuungsmaß den Interdezilbereich (IDB), d.h. die Streubreite der mittleren 80 Prozent aller Fälle.

Bitte legt die ausgefüllten Fragebögen möglichst bis Mittwoch, den 3. März, zurück in unser Fach. Wir melden uns wieder, wenn alle Bögen ausgewertet sind.

Vielen Dank für Eure Mitarbeit!

Michael und Sabine Häder



### 3. Welle

#### Verteilungsschätzungen:

##### 1. Frage (24.1)<sup>19)</sup>

Die meisten Menschen gehen ja ganz unterschiedlich mit der Zeit in ihrem Leben um und planen auch unterschiedlich. Wie gut beschreiben die Sätze auf diesen Kärtchen Deine Meinung?

Heute ist heute und morgen ist morgen.	MW	IDB	N	
überhaupt nicht	18%	15 - 25 %	18	....%
weniger gut	33%	20 - 40 %	18	....%
gut	32%	20 - 40 %	18	....%
sehr gut	17%	10 - 20 %	18	....%

##### 2. Frage (24.2.) (gleicher Fragetext)

Was ich nächste Woche machen werde, überlege ich mir dann, wenn es soweit ist.

überhaupt nicht	14%	10 - 20 %	19	....%
weniger gut	33%	25 - 40 %	19	....%
gut	35%	30 - 50 %	19	....%
sehr gut	18%	10 - 25 %	19	....%

##### 3. Frage (24.6.) (gleicher Fragetext)

Ich tue am liebsten spontan das, wozu ich gerade Lust habe.

überhaupt nicht	9%	5 - 20 %	19	....%
weniger gut	20%	10 - 30 %	19	....%
gut	44%	30 - 50 %	19	....%
sehr gut	27%	15 - 40 %	19	....%

##### 4. Frage (24.10.) (gleicher Fragetext)

Ich höre gerne Geschichten aus guten alten Zeiten.

überhaupt nicht	27%	20 - 40 %	18	....%
weniger gut	42%	30 - 50 %	18	....%
gut	20%	12 - 25 %	18	....%
sehr gut	11%	5 - 20 %	18	....%

##### 5. Frage (24. 23.) (gleicher Fragetext)

Im Leben ist alles Zufall.

überhaupt nicht	21%	10 - 30 %	19	....%
weniger gut	37%	30 - 45 %	19	....%
gut	27%	20 - 40 %	19	....%
sehr gut	15%	10 - 20 %	19	....%

---

19) Die Zahl in den Klammern gibt jeweils an, an welcher Stelle die Frage im Originalfragebogen stand.

Hinweis: Die folgenden Fragen sind als Alternativen gedacht. Die Antworten müssen deshalb zusammen 100% ergeben.

#### 6. Frage (25.a.)

Auf diesen Karten stehen jeweils unterschiedliche Meinungen, wie man mit seinem Leben umgehen kann. Sage mir bitte, welcher Meinung Du eher zustimmst.

A1 Ich betrachte mein Leben als eine Aufgabe, für die ich da bin und für die ich alle Kräfte einsetze. Ich möchte in meinem Leben etwas leisten, auch wenn das oft schwer und mühsam ist.

MW: 39% IDB: 30 - 50% N: 20 ....%

A2 Ich möchte mein Leben genießen und mich nicht mehr abmühen als nötig. Man lebt schließlich nur einmal, die Hauptsache ist doch, daß man etwas von seinem Leben hat.

MW: 61% IDB: 50 - 70% N: 20 ....%

#### 7. Frage (25.b.) (gleicher Fragetext)

B1 Ich finde es wichtig, mein Leben so einzurichten, daß ich ein ganz anderer Mensch bin, der anders ist als alle anderen Menschen in meiner Umgebung.

MW: 47% IDB: 40 - 60% N: 20 ....%

B2 Ich finde es wichtig, mein Leben so einzurichten, daß betont wird, was ich mit anderen Menschen um mich herum gemeinsam habe und worin wir uns ähnlich sind.

MW: 53% IDB: 40 - 60% N: 20 ....%

#### 8. Frage (25. c.) (gleicher Fragetext)

C1 Ich finde es am wichtigsten, daß ich im Leben selbständig bin und wirklich selbständig meine eigenen Interessen und Ziele verfolge.

MW: 64% IDB: 50 - 75% N: 20 ....%

C2 Ich finde es am wichtigsten, daß ich im Leben und meinen Entscheidungen die Interessen und Ziele anderer Menschen berücksichtige.

MW: 36% IDB: 25 - 50% N: 20 ....%

#### 9. Frage (25. d.) (gleicher Fragetext)

D1 Ich richte mein Leben so ein, daß die Dinge, die ich tue, mir sogleich und direkt etwas bringen und daß ich unmittelbar einen Nutzen sehe oder Spaß daran habe.

MW: 50% IDB: 40 - 65% N: 20 ....%

D2 Ich richte mein Leben so ein, daß die Dinge, die ich tue, sich langfristig auszahlen, daß ich später im Leben einmal etwas davon habe und auf lange Frist die Früchte meines Tuns ernten kann.

MW: 50% IDB: 35 - 60% N: 20 ....%

## Mittelwertschätzungen

### 10. Frage (27.)

Auf dieser Liste stehen einige Dinge, die wichtig als Werte für das eigene Leben sein können, was man anstrebt und wie man leben möchte. Sage Du mir bitte bei jedem, wie wichtig es Dir ist für Dein Leben.

1 = wäre nicht wichtig und 7 = wäre äußerst wichtig

#### 01. Innere Harmonie (in Frieden mit mir selbst)

MW: 5,3      IDB: 5,0 - 6,0      N: 20      geschätzter Mittelwert: ...

#### 02. Soziale Macht (Kontrolle über andere, Dominanz)

MW: 3,5      IDB: 2,0 - 4,5      N: 20      geschätzter Mittelwert: ...

#### 03. Freiheit (Freiheit des Handelns und des Denkens)

MW: 6,2      IDB: 6,0 - 6,5      N: 20      geschätzter Mittelwert: ...

#### 04. Soziale Ordnung (Stabilität der Gesellschaft)

MW: 4,4      IDB: 3,0 - 5,2      N: 20      geschätzter Mittelwert: ...

#### 05. Ein anregendes Leben (anregende Erfahrungen)

MW: 5,7      IDB: 5,0 - 6,1      N: 20      geschätzter Mittelwert: ...

#### 06. Höflichkeit (gute Umgangsformen)

MW: 3,5      IDB: 2,0 - 4,5      N: 20      geschätzter Mittelwert: ...

#### 07. Reichtum (materieller Besitz, Geld)

MW: 4,6      IDB: 4,0 - 5,5      N: 20      geschätzter Mittelwert: ...

#### 08. Nationale Sicherheit (Schutz meiner Nation gegen Feinde)

MW: 3,7      IDB: 3,0 - 4,5      N: 20      geschätzter Mittelwert: ...

#### 09. Kreativität (Originalität, Phantasie)

MW: 5,1      IDB: 4,5 - 6,0      N: 20      geschätzter Mittelwert: ...

#### 10. Eine Welt in Frieden (frei von Krieg und Konflikt)

MW: 6,1      IDB: 5,0 - 7,0      N: 20      geschätzter Mittelwert: ...

11. Achtung vor Tradition (Erhaltung ehrwürdiger Sitten)

MW: 2,6 IDB: 2,0 - 3,2 N: 20

geschätzter Mittelwert: ...

12. Loslösung (von weltlichen Belangen)

MW: 2,7 IDB: 2,0 - 3,5 N: 20

geschätzter Mittelwert: ...

13. Familiäre Sicherheit (Sicherheit für die geliebten Personen)

MW: 5,3 IDB: 4,5 - 6,0 N: 20

geschätzter Mittelwert: ...

14. Einheit mit der Natur (Einpassung in die Natur)

MW: 5,1 IDB: 4,0 - 6,0 N: 20

geschätzter Mittelwert: ...

15. Ein abwechslungsreiches Leben (erfüllt mit Herausforderungen, Neuem und Veränderungen)

MW: 5,3 IDB: 4,0 - 6,0 N: 20

geschätzter Mittelwert: ...

16. Autorität (ein Recht zu führen und zu bestimmen)

MW: 3,6 IDB: 2,5 - 4,5 N: 20

geschätzter Mittelwert: ...

17. Wahre Freundschaft (enge unterstützende Freunde)

MW: 5,7 IDB: 5,0 - 6,1 N: 20

geschätzter Mittelwert: ...

18. Eine Welt der Schönheit (Schönheit der Natur und Künste)

MW: 3,6 IDB: 3,0 - 4,0 N: 20

geschätzter Mittelwert: ...

Jetzt geht es wieder um Eure Meinung! Wir bitten Euch auch dieses Mal, die folgende Frage zu beantworten:

Wie sicher wart ihr Euch bei der Beantwortung der Fragen? Wenn ihr Euch völlig unsicher wart, kreuzt bitte das Kästchen ganz links an, wenn ihr Euch völlig sicher wart, kreuzt das Kästchen ganz rechts an. Mit den Kästchen zwischen den beiden Endpunkten könnt Ihr abstufen.

völlig  
unsicher

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

völlig  
sicher

Nochmals vielen Dank für Eure Mitarbeit.

## Anlage 4

nach Abschluß der letzten Welle wollen wir noch einige Informationen für unsere Auswertung sammeln. Bitte beantworte nun am Ende unseres Tests noch die Fragen im folgenden Bogen, in dem es um Deine persönlichen Eindrücke geht.

Bitte entscheide Dich zunächst jeweils für eine der vorgegebenen Antwortstufen!

### 1. Wie verständlich war Dir das Anliegen unseres Tests?

völlig  
verständlich

--	--	--	--	--	--	--

überhaupt  
nicht  
verständlich

### 2. Wie intensiv hast Du Dich bisher mit Fragen des Zeitbewußtseins - wie sie in unserem Test vorkamen - beschäftigt?

sehr  
intensiv

--	--	--	--	--	--	--

überhaupt  
nicht

### 3. Wie intensiv hast Du Dich bisher mit Wichtigkeitsurteilen beschäftigt?

sehr  
intensiv

--	--	--	--	--	--	--

überhaupt  
nicht

### 4. Wie stark hast Du Dich bei Deiner ersten Schätzung von folgendem leiten lassen?

Ich habe mich leiten lassen...

#### a) von meinen Erfahrungen mit ähnlichen Fragestellungen:

sehr  
stark

--	--	--	--	--	--	--

gar  
nicht

#### b) von meinen Erfahrungen mit statistischen Verteilungsformen:

sehr  
stark

--	--	--	--	--	--	--

gar  
nicht

#### c) spontan von meinem Gefühl:

sehr  
stark

--	--	--	--	--	--	--

gar  
nicht

d) von meinen Erfahrungen aus dem Umgang mit Jugendlichen:

sehr  
stark

--	--	--	--	--	--	--

gar  
nicht

e) von etwas anderem, nämlich:

.....

5. Wovon hast Du Dich bei Deinen Schätzungen in den folgenden Wellen leiten lassen?

a) Ich habe (nochmals) über den Inhalt nachgedacht.

trifft  
völlig  
zu

--	--	--	--	--	--	--

trifft  
nicht  
zu

b) Ich habe über die rückgemeldeten statistischen Verteilungsformen nachgedacht.

trifft  
völlig  
zu

--	--	--	--	--	--	--

trifft  
nicht  
zu

c) Ich habe über den Abstand meiner Schätzung zum Gruppenmittel nachgedacht.

trifft  
völlig  
zu

--	--	--	--	--	--	--

trifft  
nicht  
zu

d) Ich war bemüht, bei den verschiedenen Wellen möglichst ein konstantes Urteil abzugeben.

trifft  
völlig  
zu

--	--	--	--	--	--	--

trifft  
nicht  
zu

e) von etwas anderem, nämlich:

.....

6. Für wie kompetent hältst Du beim ZUMA rekrutierte Experten für die Schätzung solcher Befragungsergebnisse?

sehr  
kompe-  
tent

--	--	--	--	--	--	--

nicht  
kompe-  
tent

**7. Wieviel Kontakte hast Du zu Jugendlichen in dem Alter (13 bis 29 Jahre), nach deren Meinungen wir gefragt haben?**

sehr  
viele

--	--	--	--	--	--	--

gar  
keine

**8. In welchem Fach hast Du Deinen Abschluß bzw. falls Du HIWI bei ZUMA bist, was ist Deine gegenwärtige Studienrichtung?**

.....

**9. Seit wievielen Jahren befaßt Du Dich mit empirisch-sozialwissenschaftlichen Problemen?**

Seit .....Jahren.

**10. Wie lange hast Du etwa für das Ausfüllen der Fragebögen benötigt?**

In der ersten Welle etwa:

In der zweiten Welle etwa:

In der dritten Welle etwa:

.... Minuten

....Minuten

....Minuten

**11. Gibt es noch etwas, das wir über diesen Test von Dir wissen sollten? Wir interessieren uns für alles, was Dir beim Ausfüllen aufgefallen ist!**

Zuächst nochmals vielen Dank für Deine Mitarbeit. Die Ergebnisse dieses Tests werden wir, wenn die Auswertung abgeschlossen ist, beim ZUMA vorstellen.





## ZUMA-Arbeitsberichte

- 80/15 Gerhard Arminger, Willibald Nagl, Karl F. Schuessler  
Methoden der Analyse zeitbezogener Daten. Vortragsskripten der ZUMA  
Arbeitstagung vom 25.9. bis 5. 10. 79
- 81/07 Erika Brückner, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer, Peter Schmidt  
Methodenbericht zum "ALLBUS 1980"
- 81/19 Manfred Küchler, Thomas P. Wilson, Don H. Zimmerman  
Integration von qualitativen und quantitativen Forschungsansätzen
- 82/03 Gerhard Arminger, Horst Busse, Manfred Küchler  
Verallgemeinerte Lineare Modelle in der empirischen Sozialforschung
- 82/08 Glenn R. Carroll  
Dynamic analysis of discrete dependent variables: A didactic essay
- 82/09 Manfred Küchler  
Zur Messung der Stabilität von Wählerpotentialen
- 82/10 Manfred Küchler  
Zur Konstanz der Recallfrage
- 82/12 Rolf Porst  
"ALLBUS 1982" - Systematische Variablenübersicht und erste Ansätze zu einer  
Kritik des Fragenprogramms
- 82/13 Peter Ph. Mohler  
SAR - Simple AND Retrieval mit dem Siemens-EDT-  
Textmanipulationsprogramm
- 82/14 Cornelia Krauth  
Vergleichsstudien zum "ALLBUS 1980"
- 82/21 Werner Hagstotz, Hans-Peter Kirschner, Rolf Porst, Peter Prüfer  
Methodenbericht zum "ALLBUS 1982"
- 83/09 Bernd Wegener  
Two approaches to the analysis of judgments of prestige: Interindividual  
differences and the general scale
- 83/11 Rolf Porst  
Synopsis der ALLBUS-Variablen. Die Systematik des ALLBUS-  
Fragenprogramms und ihre inhaltliche Ausgestaltung im ALLBUS 1980 und  
ALLBUS 1982
- 84/01 Manfred Küchler, Peter Ph. Mohler  
Qualshop (ZUMA-Arbeitstagung zum "Datenmanagement bei qualitativen  
Erhebungsverfahren") - Sammlung von Arbeitspapieren und -berichten,  
Teil I + II
- 84/02 Bernd Wegener  
Gibt es Sozialprestige? Konstruktion und Validität der Magnitude-Prestige-  
Skala
- 84/03 Peter Prüfer, Margrit Rexroth  
Erfahrungen mit einer Technik zur Bewertung von Interviewerverhalten
- 84/04 Frank Faulbaum  
Ergebnisse der Methodenstudie zur internationalen Vergleichbarkeit von  
Einstellungsskalen in der Allgemeinen Bevölkerungsumfrage der  
Sozialwissenschaften (ALLBUS) 1982

- 84/05 Jürgen Hoffmeyer-Zlotnik  
Wohnquartiersbeschreibung. Ein Instrument zur Bestimmung des sozialen Status von Zielhaushalten
- 84/07 Gabriele Hippler, Hans-Jürgen Hippler  
Reducing Refusal Rates in the Case of Threatening Questions: The "Door-in-the-Face" Technique
- 85/01 Hartmut Esser  
Befragtenverhalten als "rationales Handeln" - Zur Erklärung von Antwortverzerrungen in Interviews
- 85/03 Rolf Porst, Peter Prüfer, Michael Wiedenbeck, Klaus Zeifang  
Methodenbericht zum "ALLBUS 1984"
- 86/01 Dagmar Krebs  
Zur Konstruktion von Einstellungsskalen im interkulturellen Vergleich
- 86/02 Hartmut Esser  
Können Befragte lügen? Zum Konzept des "wahren Wertes" im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung
- 86/03 Bernd Wegener  
Prestige and Status as Function of Unit Size
- 86/04 Frank Faulbaum  
Very Soft Modeling: The Logical Specification and Analysis of Complex Process Explanations with Arbitrary Degrees of Underidentification and Variables of Arbitrary Aggregation and Measurement Levels
- 86/05 Peter Prüfer, Margrit Rexroth (Übersetzung Dorothy Duncan)  
On the Use of the Interaction Coding Technique
- 86/06 Hans-Peter Kirschner  
Zur Kessler-Greenberg-Zerlegung der Varianz der Meßdifferenz zwischen zwei Meßzeitpunkten einer Panel-Befragung
- 86/07 Georg Erdmann  
Ansätze zur Abbildung sozialer Systeme mittels nicht-linearer dynamischer Modelle
- 86/09 Heiner Ritter  
Einige Ergebnisse von Vergleichstests zwischen den PC- und Mainframe-Versionen von SPSS und SAS
- 86/11 Günter Rothe  
Bootstrap in generalisierten linearen Modellen
- 87/01 Klaus Zeifang  
Die Test-Retest-Studie zum ALLBUS 1984 - Tabellenband
- 87/02 Klaus Zeifang  
Die Test-Retest-Studie zum ALLBUS 1984 - Abschlußbericht
- 87/04 Barbara Erbslöh, Michael Wiedenbeck  
Methodenbericht zum "ALLBUS 1986"
- 87/05 Norbert Schwarz, Julia Bienias  
What Mediates the Impact of Response Alternatives on Behavioral Reports?
- 87/06 Norbert Schwarz, Fritz Strack, Gesine Müller, Brigitte Chassein  
The Range of Response Alternatives may determine the Meaning of the Question: Further Evidence on Informative Functions of Response Alternatives
- 87/07 Fritz Strack, Leonard L. Martin, Norbert Schwarz  
The Context Paradox in Attitude Surveys: Assimilation or Contrast?

- 87/08 Gudmund R. Iversen  
Introduction to Contextual Analysis
- 87/09 Seymour Sudman, Norbert Schwarz  
Contributions of Cognitive Psychology to Data Collection in Marketing Research
- 87/10 Norbert Schwarz, Fritz Strack, Denis Hilton, Gabi Naderer  
Base-Rates, Representativeness, and the Logic of Conversation
- 87/11 George F. Bishop, Hans-Jürgen Hippler, Norbert Schwarz, Fritz Strack  
A Comparison of Response Effects in Self-Administered and Telephone Surveys
- 87/12 Norbert Schwarz  
Stimmung als Information. Zum Einfluß von Stimmungen und Emotionen auf evaluative Urteile
- 88/01 Antje Nebel, Fritz Strack, Norbert Schwarz  
Tests als Treatment: Wie die psychologische Messung ihren Gegenstand verändert
- 88/02 Gerd Bohner, Herbert Bless, Norbert Schwarz, Fritz Strack  
What Triggers Causal Attributions? The Impact of Valence and Subjective Probability
- 88/03 Norbert Schwarz, Fritz Strack  
The Survey Interview and the Logic of Conversation: Implications for Questionnaire Construction
- 88/04 Hans-Jürgen Hippler, Norbert Schwarz  
"No Opinion"-Filters: A Cognitive Perspective
- 88/05 Norbert Schwarz, Fritz Strack  
Evaluating One's Life: A Judgment of Subjective Well-Being
- 88/06 Norbert Schwarz, Herbert Bless, Gerd Bohner, Uwe Harlacher, Margit Kellenbenz  
Response Scales as Frames of Reference: The Impact of Frequency Range on Diagnostic Judgments
- 88/07 Michael Braun  
Allbus-Bibliographie (7. Fassung, Stand: 30.6.88)
- 88/08 Günter Rothe  
Ein Ansatz zur Konstruktion inferenzstatistisch verwertbarer Indices
- 88/09 Ute Hauck, Reiner Trometer  
Methodenbericht International Social Survey Program - ISSP 1987
- 88/10 Norbert Schwarz  
Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction
- 88/11 Norbert Schwarz, B. Scheuring (sub.)  
Judgments of relationship satisfaction: Inter- and intraindividual comparison strategies as a function of questionnaire structure
- 88/12 Rolf Porst, Michael Schneid  
Ausfälle und Verweigerungen bei Panelbefragungen - Ein Beispiel -
- 88/13 Cornelia Züll  
SPSS-X. Anmerkungen zur Siemens BS2000 Version
- 88/14 Michael Schneid  
Datenerhebung am PC - Vergleich der Interviewprogramme "internerv+" und "THIS"

- 88/15 Norbert Schwarz, Bettina Scheuring  
Die Vergleichsrichtung bestimmt das Ergebnis von Vergleichsprozessen: Ist -  
Idealdiskrepanzen in der Partnerwahrnehmung
- 88/16 Norbert Schwarz, Bettina Scheuring  
Die Vergleichsrichtung bestimmt das Ergebnis von Vergleichsprozessen: Ist-  
Idealdiskrepanzen in der Beziehungsbeurteilung
- 89/01 Norbert Schwarz, George F. Bishop, Hans-J. Hippler, Fritz Strack  
Psychological Sources of Response Effects in Self-Administered And Telephone  
Surveys
- 89/02 Michael Braun, Reiner Trometer, Michael Wiedenbeck  
Methodenbericht. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften -  
ALLBUS 1988
- 89/03 Norbert Schwarz  
Feelings as Information: Informational and Motivational Functions of Affective  
States
- 89/04 Günter Rothe  
Jackknife and Bootstrap: Resampling-Verfahren zur Genauigkeitsschätzung von  
Parameterschätzungen
- 89/05 Herbert Bless, Gerd Bohner, Norbert Schwarz und Fritz Strack  
Happy and Mindless? Moods and the Processing of Persuasive Communications
- 89/06 Gerd Bohner, Norbert Schwarz und Stefan E. Hormuth  
Die Stimmungs-Skala: Eine deutsche Version des "Mood Survey" von  
Underwood und Froming
- 89/07 Ulrich Mueller  
Evolutionary Fundamentals of Social Inequality, Dominance and Cooperation
- 89/08 Robert Huckfeldt  
Noncompliance and the Limits of Coercion: The Problematic Enforcement of  
Unpopular Laws
- 89/09 Peter Ph. Mohler, Katja Frehsen und Ute Hauck  
CUI - Computerunterstützte Inhaltsanalyse. Grundzüge und  
Auswahlbibliographie zu neueren Anwendungen
- 89/10 Cornelia Züll, Peter Ph. Mohler  
Der General Inquirer III - Ein Dinosaurier für die historische Forschung
- 89/11 Fritz Strack, Norbert Schwarz, Brigitte Chassein, Dieter Kern, Dirk Wagner  
The Salience of Comparison Standards and the Activation of Social Norms:  
Consequences for Judgments of Happiness and their Communication
- 89/12 Jutta Kreiselmaier, Rolf Porst  
Methodische Probleme bei der Durchführung telefonischer Befragungen:  
Stichprobenziehung und Ermittlung von Zielpersonen, Ausschöpfung und  
Nonresponse, Qualität der Daten
- 89/13 Rainer Mathes  
Modulsystem und Netzwerktechnik. Neuere inhaltsanalytische Verfahren zur  
Analyse von Kommunikationsinhalten
- 89/14 Jutta Kreiselmeier, Peter Prüfer, Margrit Rexroth  
Der Interviewer im Pretest. Evaluation der Interviewerleistung und Entwurf  
eines neuen Pretestkonzepts. April 1989
- 89/15 Henrik Tham  
Crime as a Social Indicator

- 89/16 Ulrich Mueller  
Expanding the Theoretical and Methodological Framework of Social Dilemma Research
- 89/17 Hans-J. Hippler, Norbert Schwarz, Elisabeth Noelle-Neumann  
Response Order Effects in Dichotomous Questions: The Impact of Administration Mode
- 89/18 Norbert Schwarz, Hans-J. Hippler, Elisabeth Noelle-Neumann, Thomas Münkler  
Response Order Effects in Long Lists: Primacy, Recency, and Asymmetric Contrast Effects
- 89/19 Wolfgang Meyer  
Umweltberichterstattung in der Bundesrepublik Deutschland
- 89/20 Michael Braun, Reiner Trometer  
ALLBUS Bibliographie (8. Fassung, Stand: 30.6.1989)
- 89/21 Günter Rothe  
Gewichtungen zur Anpassung an Statusvariablen. Eine Untersuchung am ALLBUS 1986
- 89/22 Norbert Schwarz, Thomas Münkler, Hans-J. Hippler  
What determines a "Perspective"? Contrast Effects as a Function of the Dimension Tapped by Preceding Questions
- 89/23 Norbert Schwarz, Andreas Bayer  
Variationen der Fragenreihenfolge als Instrument der Kausalitätsprüfung: Eine Untersuchung zur Neutralisationstheorie devianten Verhaltens
- 90/01 Norbert Schwarz, Andreas Bayer  
Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis
- 90/02 Norbert Schwarz, Fritz Strack, Hans-J. Hippler, George Bishop  
The Impact of Administration Mode on Response Effects in Survey Measurement
- 90/03 Norbert Schwarz, Herbert Bless, Gerd Böhner  
Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications
- 90/04 Michael Braun, Reiner Trometer  
ALLBUS-Bibliographie 90
- 90/05 Norbert Schwarz, Fritz Strack  
Context Effects in Attitude Surveys: Applying Cognitive Theory to Social Research
- 90/06 Norbert Schwarz, Herbert Bless, Fritz Strack, Gisela Klumpp, Annette Simons  
Ease of Retrieval as Information: Another Look at the Availability Heuristic
- 90/07 Norbert Schwarz, Fritz Strack, Hans-J. Hippler  
Kognitionspsychologie und Umfrageforschung: Themen und Befunde eines interdisziplinären Forschungsgebietes
- 90/08 Norbert Schwarz, Hans-J. Hippler  
Response Alternatives: The Impact of their Choice and Presentation Order
- 90/09 Achim Koch  
Externe Vergleichsdaten zum ALLBUS 1984, 1986, 1988
- 90/10 Norbert Schwarz, Bärbel Knäuper, Hans-J. Hippler, Elisabeth Noelle-Neumann, Leslie Clark  
Rating Scales: Numeric Values May Change the Meaning of Scale Labels

91/01	Denis J. Hilton Conversational Inference and Rational Judgment
91/02	Denis J. Hilton A Conversational Model of Causal Explanation
91/03	Mood Effects on Interpersonal Preferences: Evidence for Motivated Processing Strategies
91/04	Joseph P. Forgas Affective Influences on Interpersonal Perception
91/05	Norbert Schwarz, Herbert Bless Constructing Reality and Its Alternatives: An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgment
91/06	Herbert Bless, Roland F. Fellhauer, Gerd Bohner, Norbert Schwarz Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben
91/07	Norbert Schwarz, Bärbel Knäuper, E. Tory Higgins Der Einfluß von Rangordnungsaufgaben auf nachfolgende Denkprozesse: Zur Aktivierung prozeduraler Sets
91/08	Bettina Scheuring, Norbert Schwarz Selbstberichtete Verhaltens- und Symptommhäufigkeiten: Was Befragte aus Antwortvorgaben des Fragebogens lernen
91/09	Norbert Schwarz, Herbert Bless Scandals and the Public's Trust in Politicians: Assimilation and Contrast Effects
91/10	Rolf Porst Ausfälle und Verweigerungen bei einer telefonischen Befragung
91/11	Uwe Blien, Heike Wirth, Michael Müller Identification risk for microdata stemming from official statistics
91/12	Petra Beckmann Methodological Report ISSP 1989
91/13	Martina Wasmer, Achim Koch, Michael Wiedenbeck Methodenbericht zur "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (Allbus) 1990
91/14	Uwe Blien, Oded Löwenbein Einkommensanalysen auf der Grundlage amtlicher Daten und Umfragedaten: Ergebnisse zur betrieblichen Seniorität und Arbeitslosigkeit
91/15	Petra Beckmann, Peter Mohler, Rolf Uher ISSP Basic Information on the ISSP Data Collection 1985 -1994
91/16	Norbert Schwarz In welcher Reihenfolge fragen? Kontexteffekte in standardisierten Befragungen
91/17	Ellen D. Riggle, Victor C. Ottati, Robert S. Wyer, Jr., James Kuklinski, Norbert Schwarz Bases of Political Judgments: The Role of Stereotypic and Non-stereotypic Information
91/18	Dagmar Krebs Was ist sozial erwünscht? Der Grad sozialer Erwünschtheit von Einstellungseitems
91/19	Michael Braun, Reiner Trometer ALLBUS-Bibliographie
91/20	Michael Schneid Einsatz computergestützter Befragungssysteme in der Bundesrepublik Deutschland

- 91/21 Rolf Porst, Michael Schneid  
Software-Anforderungen an computergestützte Befragungssysteme
- 91/22 Ulrich Mueller  
The Reproductive Success of the Elites in Germany, Great Britain, Japan and the USA during the 19th and 20th Century
- 92/01 P.H. Hartmann, B. Schimpl-Neimanns  
Zur Repräsentativität sozio-demographischer Merkmale des ALLBUS -  
multivariate Analysen zum Mittelschichtbias der Umfrageforschung
- 92/02 Gerd Bohner, Kimberly Crow, Hans-Peter Erb, Norbert Schwarz  
Affect and Persuasion: Mood Effects on the Processing of Message Content  
and Context Cues and on Subsequent Behavior
- 92/03 Herbert Bless, Gerd Bohner, Traudel Hild, Norbert Schwarz  
Asking Difficult Questions: Task Complexity Increases the Impact of Response  
Alternatives
- 92/04 Wolfgang Bandilla, Siegfried Gabler, Michael Wiedenbeck  
Methodenbericht zum DFG-Projekt Allbus Baseline-Studie
- 92/05 Frank Faulbaum  
Von der Variablenanalyse zur Evaluation von Handlungs- und  
Prozeßzusammenhängen
- 92/06 Ingwer Borg  
Überlegungen und Untersuchungen zur Messung der subjektiven Unsicherheit  
der Arbeitsstelle
- 92/07 Ingwer Borg, Michael Braun  
Arbeitsethik und Arbeitsinvolvement als Moderatoren der psychologischen  
Auswirkungen von Arbeitsunsicherheit
- 92/08 Eleanor Singer, Hans-Jürgen Hippler, Norbert Schwarz  
Confidentiality Assurances in Surveys: Reassurance or Threat?
- 92/09 Herbert Bless, Diane M. Mackie, Norbert Schwarz  
Mood Effects on Attitude Judgments: The Independent Effects of Mood before  
and after Message Elaboration
- 92/10 Ulrich Mueller, Carola Schmid  
Ehehäufigkeit und Fruchtbarkeit weiblicher Mitglieder der deutschen Elite
- 92/11 Herbert Bless, Fritz Strack, Norbert Schwarz  
The Informative Functions of Research Procedures: Bias and the Logic of  
Conversation
- 92/12 Norbert Schwarz, Herbert Bless, Michaela Wänke  
Subjective Assessment and Evaluation of Change: Lessons from Social  
Cognition Research
- 92/13 Norbert Schwarz, Hans-J. Hippler  
Buffer Items: When Do They Buffer and When Don't They?
- 92/14 Hans-J. Hippler, Norbert Schwarz  
The Impact of Administration Modes on Response Effects in Surveys
- 92/15 Michaela Wänke, Norbert Schwarz  
Comparative Judgments: How the Direction of Comparison Determines the  
Answer
- 92/16 Michael Braun, Reiner Trometer  
ALLBUS-Bibliographie (11. Fassung, Stand: 30.6.1992)
- 92/17 Anke Nau, Rolf Porst  
Projektbericht ZUMA-Methodenpanel Teil 1: Konzeption, Vorbereitung und  
Durchführung



- 92/18 Michael Schneid  
Handbuch ZUMA-Feldsteuerungsprogramm
- 92/19 Paul Lüttinger, Bernhard Schimpl-Neimanns  
Amtliche Bildungsstatistik und empirische Sozialforschung
- 92/20 Rolf Porst, Michael Schneid  
Fragebogenschreiben mit Microsoft WORD 5.0: Druckformate, Textbausteine und Makros zur effizienten Gestaltung von Fragebogen
- 92/21 Michael Braun, Jacqueline Scott, Duane F. Alwin  
Economic Necessity or Self-Actualization? Attitudes toward Women's Labor-force Participation in the East and West
- 92/22 Duane F. Alwin, Michael Braun, Jacqueline Scott  
The Separation of Work and the Family: Attitudes Towards Women's Labour-Force Participation in Germany, Great Britain, and the United States
- 92/23 Michaela Wänke, Norbert Schwarz, Herbert Bless  
The Availability Heuristic Revisited: Experienced Ease of Retrieval in Mundane Frequency Estimates
- 93/01 Michael Braun, Carmen Eilinghoff, Siegfried Gabler, Michael Wiedenbeck  
Methodenbericht zur Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (Allbus) 1992
- 93/02 Michael Häder, Sabine Novossadeck  
Einstellungsdynamik in Ostdeutschland. Auswertung von Wiederholungsuntersuchungen zur individuellen Bewältigung des sozialen Wandels
- 93/03 Norbert Schwarz  
Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation
- 93/04 Norbert Schwarz, Michaela Wänke, Herbert Bless  
Subjective Assessments and Evaluations of Change: Some Lessons from Social Cognition Research
- 93/05 Norbert Schwarz  
Context Effects in Attitude Measurement
- 93/06 Reiner Trometer  
Die Operationalisierung des Klassenschemas nach Goldthorpe im ALLBUS
- 93/07 Carola Schmid  
Der Zugang zu den Daten der Demographie
- 93/08 Michael Schneid  
Fragebogenschreiben mit Word für Windows
- 93/09 Achim Koch  
Die Nutzung demographischer Informationen in den Veröffentlichungen mit Allbus-Daten
- 93/10 Helmut Schröder  
Über den Zusammenhang zwischen Aktivitäten und Zufriedenheit: "Eine kommunale Seniorenbefragung"
- 93/11 Michael Braun, Reiner Trometer  
ALLBUS-Bibliographie, 12. Fassung, Stand: 30.9.93
- 93/12
- 93/13 Steven E. Finkel, Peter R. Schrott  
Campaign Effects on Voter Choice in the German Election of 1990



- 93/14 Jürgen Hoffmeyer-Zlotnik, Dagmar Krebs  
Subjektive Statuszuweisung; Objektive Schichtmessung
- 93/15 Dagmar Krebs  
Richtungseffekte von Itemformulierungen
- 93/16 Dagmar Krebs  
Social Desirability: The collective conscience? Judging the degree of social desirability in attitude items
- 93/17 Bernhard Krüger, Heiner Ritter, Cornelia Züll  
SPSS Einsatz auf unterschiedlichen Plattformen in einem Netzwerk: Daten und Ergebnisaustausch
- 94/01 Jürgen H.P. Hoffmeyer-Zlotnik, Michael Wiedenbeck  
Überlegungen zu Sampling, Qualitätsprüfung und Auswertung von Daten aus Teilpopulationen
- 94/02 Michael Häder, Sabine Häder  
Die Grundlagen der Delphi-Methode: - Ein Literaturbericht -
- 94/03 Sabine Häder  
Auswahlverfahren bei Telefonumfragen
- 94/04 Peter Prüfer, Margrit Rexroth  
Ein Verfahren zur Erfassung von Erhebungsproblemen bei Interviews der Hauptstudie